

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2003, 13:49:28 ; Search time 73 seconds
(without alignments)
1038.625 Million cell updates/sec

Title: US-09-805-337A-2

Perfect score: 3183

Sequence: 1 MLLLFVILLISWTVGGEG.....SSPFRALCOEGKFEYPICE 569

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3183	100.0	569	AAE23195	Human factor H-rel
2	1691	53.1	327	AB80571	Human sbg614126com
3	1599	50.2	364	AB80570	Human sbg614126com
4	1492	46.9	267	ABG56700	Human novel polype
5	1303	40.9	678	ABP41931	Human ovarian anti
6	1051.5	33.0	342	AA843738	Human cancer assoc
7	1001.5	31.5	578	AA09065	Human complement f
8	756.5	23.8	660	9 AAP82921	B subunit of human
9	708.5	22.3	582	22 AAU02929	Angiotensin conver
10	562	17.7	290	18 AAW39156	Human partial comp

11	466	14.6	229	18 AAW39158	Clone PRBS3FH2910
12	432.5	13.6	196	18 AAW39159	Clone PRBS3FH2910
13	408	12.8	1087	12 AAR11139	B lymphocyte membr
14	395	12.4	979	22 AAG66398	Receptor 222. Uni
15	395	12.4	1066	22 ABP37987	Human GS96663 prot
16	395	12.4	1139	22 ABH73355	Novel central nerv
17	395	12.4	1139	22 ABB10326	Human CDNA SEQ ID
18	395	12.4	1139	22 AAU18126	Novel human uterin
19	395	12.4	1139	22 AAU16963	Human novel secret
20	395	12.4	1139	22 AAU19902	Novel human calciu
21	395	12.4	1251	23 AAU74829	Human REPTR 12 pro
22	394	12.4	3571	23 AAU20146	Human C3b/C4b comp
23	389.5	12.2	330	22 AAU02952	Angiotensin conver
24	389	12.2	2050	23 AAG68264	Human POLY11 prote
25	385	12.1	1537	12 AAR11982	Partial human comp
26	385	12.1	1929	22 ABG00103	Novel human diagno
27	385	12.1	1930	19 AAU45899	Human complement r
28	385	12.1	2039	12 AAR11810	Human complement t
29	385	12.1	2039	20 AAU55751	Human C3b/C4b rece
30	385	12.1	2039	22 ABG00287	Novel human diagno
31	385	12.1	2039	23 AAW50797	Human C3B/C4B rece
32	385	12.1	2044	22 AAW39224	Human polypeptide
33	385	12.1	2317	10 AAP92219	CR1 protein. Homo
34	384.5	12.1	3594	23 AAU20147	Mouse C3b/C4b comp
35	383.5	12.0	841	23 ABP51418	Human MDR1 SEQ ID
36	382	12.0	1497	22 AAW93954	Human polypeptide,
37	380.5	12.0	778	19 AAW73147	Amino acid sequenc
38	380.5	12.0	2044	22 ABB11782	Human CR1 protein
39	380.5	12.0	2044	22 AAW41010	Human polypeptide
40	379.5	11.9	645	21 AAB53125	Macaca mulatta rha
41	376	11.8	2039	14 AAR36743	CR1. Homo sapiens
42	370	11.6	3069	23 AAE20787	Human C3b/C4b comp
43	370	11.6	3069	23 AAE20900	Human C3b/C4b comp
44	370	11.6	3100	23 AAE20789	Human C3b/C4b comp
45	370	11.6	3100	23 AAE20901	Human C3b/C4b comp

ALIGNMENTS

RESULT 1	
AAE23195	
ID	AAE23195 standard; Protein; 569 AA.
XX	
AC	AAE23195;
XX	
DT	21-AUG-2002 (first entry)
XX	
DE	Human factor H-related protein (FHR-5).
XX	
KW	Human; factor H-related protein; FHR-5; detection; C5b-9 complement;
KW	immunohistochemical; diagnostic; immunological; biopsy; prevention.
XX	
OS	Homo sapiens.
XX	
PN	WO200168695-A2.
XX	
PD	20-SEP-2001.
XX	
PF	13-MAR-2001; 2001WO-US07868.
XX	
PR	13-MAR-2000; 2000US-0188870.
XX	
PA	(BAXT) BAXTER INT INC.
XX	
PI	Murphy BF;
XX	
DR	WPI; 2001-582437/65.
DR	N-PSDB; AAD37114.
XX	
PT	New factor-H related protein 5 that binds to complement component C3b,
PT	useful for raising antibodies suitable for detection of C5b-9
PT	complement complexes,

QY 241 FIINGPKKIOCVGEGWTLPTCEVQKTCGYIPELEYGYVQSPVPPYQHGVSVEVNCRNE 300
Db 144 -----MKTCTGYIPELEYGYVQSPVPPYQHGVSVEVNCRNE 178
QY 301 YAMIGNNMITCINGIWTLPVATHQKRCIKAGVNIKTLKLSGKEFNHNSRIRYRCS 360
Db 179 YAMIGNNMITCINGIWTLP----- 199
QY 361 DIFRYHVSVCINGKWNPEVDCTEKREQFCPPPPQIPNAQNMTTIVNYQDGEKVAVLCKEN 420
Db 200 ----- 199
QY 421 YLLPEAKEIVCKDGRWQSLPRCVSTAYCGPPPSINNGDTTSFPLSVYPPGSTVYRQCS 480
Db 200 -----CVSTAYCGPPPSINNGDTTSFPLSVYPPGSTVYRQCS 238
QY 481 FYKLOGSVTVTCRNKQKQSEPPRCIDPCVSVSENNKNNIQLKWRNDGKLYAKTGDAVEFQ 540
Db 239 FYKLOGSVTVTCRNKQKQSEPPRCIDPCVSVSENNKNNIQLKWRNDGKLYAKTGDAVEFQ 298
QY 541 CKFPHKAMISSPPFRAICQEGKFEYPICE 569
Db 299 CKFPHKAMISSPPFRAICQEGKFEYPICE 327

RESULT 3

ABB80570
ID ABB80570 standard; Protein; 364 AA.
XX AC ABB80570;
XX DT 08-OCT-2002 (first entry)
XX DE Human sbg614126complifh protein #1.
XX KW Human; secreted protein; immunosuppressive; cytostatic; nootropic;
KW neuroprotective; antitumour; vulnary; antimicrobial; ophthalmological;
KW antiparkinsonian; antirheumatic; antiatherosclerotic; dermatological;
KW hypotensive; cerebroprotective; virucide; antiinflammatory; diabetes;
KW malignant tumour; hypertension; hypotension; obesity; bulimia; anorexia;
KW asthma; manic depression; dementia; delirium; mental retardation;
KW Huntington's disease; Tourette's syndrome; schizophrenia;
KW mental disorder; sexual development disorder; blood cascade dysfunction;
KW stroke; growth disorder.
XX OS Homo sapiens.
XX PN WO200222802-A1.
XX PD 21-MAR-2002.
XX PF 13-SEP-2001; 2001WO-US28462.
XX PR 13-SEP-2000; 2000US-232455P.
XX PR 13-SEP-2000; 2000US-232463P.
XX PR 02-OCT-2000; 2000US-237293P.
XX PR 07-NOV-2000; 2000US-246269P.
XX PR 20-NOV-2000; 2000US-252049P.
XX PA (SMIK) SMITHKLINE BEECHAM CORP.
PA (SMIK) SMITHKLINE BEECHAM PLC.
PA (GLAX) GLAXO GROUP LTD.
XX PI Agarwal P, Gogswell JP, Lai Y, Martensen SA, Smith RF, Strum JC;
PI Xie Q;
XX WPI; 2002-393963/42.
XX N-PSDB; ABL57706.
XX PT Novel isolated secreted polypeptides and polynucleotides encoding them
PT useful for treating cancer, Alzheimer's disease, tumor metastasis,
PT autosomal recessive atypical hemolytic uremic syndrome, wound healing

PT disorder

PS 1: Page191-192; 246pp; English.

CC The invention relates to a novel isolated polypeptide (ABB80569-ABB80612)
CC (secreted polypeptide) which is encoded by any one of 44 polynucleotide
CC sequences (ABL57705-ABL57748) given in the specification. The
CC polypeptides have immunosuppressive, cytostatic, nootropic,
CC neuroprotective, antitumour, vulnary, antimicrobial, ophthalmological,
CC antiparkinsonian, antirheumatic, antiatherosclerotic, dermatological,
CC hypotensive, cerebroprotective, virucide, and antiinflammatory activity.
CC The polynucleotide and polypeptide are useful for treating diabetes,
CC malignant tumours, hyper- and hypotension, obesity, bulimia, anorexia,
CC Huntington's disease, Tourette's syndrome, schizophrenia, growth, mental
CC or sexual development disorders, and dysfunctions of the blood cascade
CC system including those leading to stroke.

XX SQ Sequence 364 AA;

Query Match 50.2%; Score 1599; DB 23; Length 364;
Best Local Similarity 55.8%; Pred. No. 3.5e-120;
Matches 307; Conservative 1; Mismatches 0; Indels 242; Gaps 2;
QY 20 GTLCDPFXKHGFLYDEEDYNPFQVPTGEVFFYSCYFNFSKSFTRITCTEGWSP 79
Db 57 GTLCDPFXKHGFLYDEEDYNPFQVPTGEVFFYSCYFNFSKSFTRITCTEGWSP 116
QY 80 TPKCLRMCSFPPVKNHSESSGLIHLEGDTVQIICNTGYSLONNEKNISCVRGHSTPPI 139
Db 117 TPKCLRMCSFPPVKNHSESSGLIHLEGDTVQIICNTGYSLONNEKNISCVRGHSTPPI 176
QY 140 CSFTKGECHVPFILEANVDAQPKESYKVGVDLKFSCRNLIRVSDSVQCYQFGWSPNP 199
Db 177 CSFT----- 180
QY 200 TKGQVRSCGPPPPQLSNGEVKEIRKEEYGHNEWEYDCNPNFIINGPKKIQCVGEGWTL 259
Db 181 ----- 180
QY 260 PTCVEQVTKCGYIPELEYGYVQSPVPPYQHGVSVEVNCRNEYAMIGNNMITCINGIWT 319
Db 181 -----MKTCTGYIPELEYGYVQSPVPPYQHGVSVEVNCRNEYAMIGNNMITCINGIWT 234
QY 320 PWCVATHQLKRCIKAGVNIKTLKLSGKEFNHNSRIRYRCSDFRYHVSVCINGKWNPEV 379
Db 235 PWC----- 236
QY 380 DCTEKREQFCPPPPQIPNAQNMTTIVNYQDGEKVAVLCKENYLLPEAKEIVCKDGRWQSL 439
Db 237 ----- 236
QY 440 PRCVESTAYCGPPPSINNGDTTSFPLSVYPPGSTVYRQCSFYKLOGSVTVTCRNKQSE 499
Db 237 --CVSTAYCGPPPSINNGDTTSFPLSVYPPGSTVYRQCSFYKLOGSVTVTCRNKQSE 294
QY 500 PPRCLDPCVSVSENNKNNIQLKWRNDGKLYAKTGDAVEFQCKFPHKAMISSPPFRAICQ 559
Db 295 PPRCLDPCVSVSENNKNNIQLKWRNDGKLYAKTGDAVEFQCKFPHKAMISSPPFRAICQ 354
QY 560 ESKFEYPICE 569
Db 355 ESKFEYPICE 364

RESULT 4

ABG66700
ID ABG66700 standard; Protein; 267 AA.
XX AC ABG66700;
XX DT 30-AUG-2002 (first entry)
XX

DE XX Human novel polypeptide #35.

KW Human; inflammatory condition; shock; sepsis; immune response;
 KW cancer; wound healing; central nervous system disease; haematopoiesis;
 KW peripheral nervous system disease; amyotrophic lateral sclerosis; tendon;
 KW myeloid cell disorder; lymphoid cell disorder; platelet disorder; bone;
 KW cartilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis;
 KW bone degenerative disorder; periodontal disease; reperfusion injury;
 KW lung fibrosis; liver fibrosis; autoimmune disorder; bacterial infection;
 KW allergic condition; thrombolysis; thrombosis; coagulation disorder;
 KW fungal infection.

XX Homo sapiens.

XX WO200244340-A2.

XX 06-JUN-2002.

XX 30-NOV-2001; 2001WO-US47004.

XX 30-NOV-2000; 2000US-0028952.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Wang J, Wang D;
 PI Yamazaki V, Ujwal ML, Drmanac RT;
 XX WPI; 2002-508509/54.
 DR N-PSDB; ABK94924.

XX Novel nucleic acids and polypeptides for diagnosis, treatment of
 PT inflammatory, autoimmune, nervous system, myeloid or lymphoid cell
 PT disorders, cancer and promoting wound healing -
 XX Claim 10; Page 601; 672pp; English.

XX The invention relates to human novel polynucleotides and associated
 CC polypeptides. The polynucleotides and polypeptides are useful for
 CC treating inflammatory conditions such as arthritis, nephritis, Crohn's
 CC disease, ischaemia-reperfusion injury, shock, sepsis, immune responses
 CC and cancer and for promoting wound healing. The sequences are used to
 CC induce the proliferation of neural cells and regeneration of nerve and
 CC brain tissue, and are useful for the treatment of central and peripheral
 CC nervous system diseases and neuropathies, such as Alzheimer's disease,
 CC Parkinson's disease, Huntington's disease and amyotrophic lateral
 CC sclerosis. The sequences are involved in chemotactic or chemokinetic
 CC activity, regulation of haematopoiesis, treatment of myeloid or lymphoid
 CC cell disorders and platelet disorders such as thrombocytopenia, tissue
 CC regeneration of bone, cartilage, tendon, ligament and/or nerve fibrosis,
 CC growth, tissue repair, healing of burns, incisions, ulcers, treatment of
 CC osteoporosis, osteoarthritis, bone degenerative disorders and periodontal
 CC disease. The sequences of the invention are also useful for gut
 CC protection or regeneration and treatment of lung or liver fibrosis,
 CC reperfusion injury in various tissues, immune deficiencies and disorders
 CC including severe combined immunodeficiency (SCID), bacterial or fungal
 CC infections, autoimmune disorders e.g. multiple sclerosis and myasthenia
 CC gravis, allergic conditions such as asthma, thrombolysis or thrombosis
 CC and coagulation disorders. Sequences ABG6666-ABG66758 represent human
 CC novel polypeptides of the invention.

XX Sequence 267 AA;

SQ Query Match 46.9%; Score 1492; DB 23; Length 267;
 Best Local Similarity 100.0%; Pred. No. 9.2e-112;
 Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 303 MIGNNMITCINGLWELPVCVATHQKCKKIAGVNIKTLLKSGKEFNHSRIRYCSDI 362
 DB 1 MIGNNMITCINGLWELPVCVATHQKCKKIAGVNIKTLLKSGKEFNHSRIRYCSDI 60
 QY 363 FRYRHSVCINGKWNPEVDCTEKREOFCPPPPQIPNQAQNTTTVYODGKVAVLCKENYL 422
 DB 61 FRYRHSVCINGKWNPEVDCTEKREOFCPPPPQIPNQAQNTTTVYODGKVAVLCKENYL 120

QY 423 LPEAKEIVCKGRWQSLPRCVESTAYCGPPPSINNGDTTSFPLSVYPPGSTVYRCOSFY 482
 DB 121 LPEAKEIVCKGRWQSLPRCVESTAYCGPPPSINNGDTTSFPLSVYPPGSTVYRCOSFY 180
 QY 483 KLGQSVTVTCRNKQWSEPPRCCLDPCVVVSEENMKNNTQLKWRNDGKLYAKTGDAVEPQCK 542
 DB 181 KLGQSVTVTCRNKQWSEPPRCCLDPCVVVSEENMKNNTQLKWRNDGKLYAKTGDAVEPQCK 240
 QY 543 FPHKAMTSSPPFRAICOEGKFEYPICE 569
 DB 241 FPHKAMTSSPPFRAICOEGKFEYPICE 267

RESULT 5
 ABP41931
 ID ABP41931 standard; Protein; 678 AA.
 XX AC ABP41931;
 XX DT 22-AUG-2002 (first entry)
 XX DE Human ovarian antigen HVVBI39, SEQ ID NO:3063.
 XX Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
 KW ovarian cancer; breast cancer; tumour; reproductive system disorder;
 KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
 KW PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;
 KW inflammatory condition; immune disorder; blood disorder;
 KW cardiovascular disorder; respiratory disorder; neurological disorder;
 KW gastrointestinal disorder; urinary system disorder; drug screening;
 KW gene therapy; chromosome mapping; forensic analysis;
 KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;
 KW antiinflammatory; gynaecological; reproductive.
 XX Homo sapiens.
 OS WO200200677-A1.
 XX 03-JAN-2002.
 XX 07-JUN-2001; 2001WO-US18569.
 XX 07-JUN-2000; 2000US-209467P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX Birse CE, Rosen CA;
 XX WPI; 2002-147878/19.
 DR N-PSDB; ABQ55008.
 XX Isolated nucleic acid molecules encoding novel ovarian polypeptides,
 PT useful in the prevention, treatment and diagnosis of cancer (e.g.
 PT ovarian cancer), immune disorders, cardiovascular disorders and
 PT neurological diseases -
 XX Claim 11; SEQ ID No 3063; 2922pp; English.
 XX The invention relates to 2175 novel human ovarian antigens (ABP41054-
 CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
 CC encompasses polypeptides 90% identical and polynucleotides 95% identical
 CC to the sequences of the invention. The invention additionally relates to
 CC recombinant vectors and host cells comprising human ovarian antigen
 CC polynucleotides, antibodies against human ovarian antigens, and the use
 CC of ovarian antigen polynucleotides and polypeptides in diagnosing,
 CC treating, prognosing or preventing various ovary and/or breast-related
 CC disorders. Such conditions include ovarian cancer and breast cancer, and
 CC metastatic tumours of ovarian or breast origin, reproductive system
 CC disorders (e.g., infertility, disorders of pregnancy, anovulation,
 CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
 CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
 CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and

QY 61 SPKSFWRITCTEAGSPKPKCLRMCSFFVKNHSESSGLIHLEGGDTVOICNTQYSL 120
 Db 73 SPKSFWRITCTEAGSPKPKCLRLCFFPVENHSESSGQTHLEGGDTVOICNTQYRL 132
 QY 121 QNNEKNISCVBERGWSPTTSCFTKGECHVPILEAVDAQPKKESYKVGDLKFSCKRNLI 180
 Db 133 QNNEKNISCVBERGWSPTTSCFTKGECHVPILEAVDAQPKKESYKVGDLKFSCKRNLI 180
 QY 181 RVGSDSVQCYOFQWSPNFPKCKQVRSCGPPPLNSGEVKEIRKEEYGHNEVVEYDCNPN 240
 Db 157 -----TSCVNPPTVQNAIXSRMSKYPGSEVRVY----- 186
 QY 241 FIINGPKKIQCVGENTTLCVEQVKTCGYIPELEYGYQVSPVPYQHOVSVEVNCRNE 300
 Db 187 -----XCRSP 191
 QY 301 YAMIGNMITCINGIWTPLMPCVATHQKRCCKIAGVNIKTLLKLSGKEFNHNSRIYRCS 360
 Db 192 YEMFGDEVMCLNGNTE----- 209
 QY 361 DIFRYHSVCINGKWNDEVDCTEKREQFCPPPIPAQNMTTNNYODGEKVAVLCKEN 420
 Db 210 -----PQO----- 212
 QY 421 YLLPEAKEIYCKDGRWQSLPRVSTAYCGPPPSINNGDTTSFPLSVYPPGTVTRCQS 480
 Db 213 -----CKD-----STGCKGPPPIDNGDITSFPLSVIAPASSVEYQCQN 251
 QY 481 FYKLOGSVTVTCRNKQWSEPPRCCLDPCVSEENKNNIOLKWRNDGKLYAKTGDAVEFQ 540
 Db 252 LYQLEGNKRITCRNGOWSEPPKCLHPCVISREIMENYIALRTAKQLYXRTGESXEFV 311
 QY 541 CKPPHKAMISPPFRAICQEGKEEYPTC 568
 Db 312 CKRGYLSRSRSHLTRTCTMDGKLEYPTC 339

RESULT 7

AA09065
 ID AAY09065 standard; Protein; 578 AA.
 AC AAY09065;
 XX AAY09065;
 DT 06-JUL-1999 (first entry)
 DE Human complement factor H homolog protein.
 KW Human complement factor H; immunological mechanism; complement reaction;
 KW gene therapy; immune stimulation; haematopoiesis regulation; chemotactic;
 KW tissue growth activity; anti-inflammatory; tumour inhibition;
 KW secretory signal.
 XX Homo sapiens.
 OS Homo sapiens.
 XX W09918200-A1.
 PN W09918200-A1.
 XX 15-APR-1999.
 PD 15-APR-1999.
 XX 02-OCT-1998; 98WO-JP04448.
 PF 02-OCT-1998; 98WO-JP04448.
 XX 06-OCT-1997; 97JP-0272837.
 PR 06-OCT-1997; 97JP-0272837.
 XX (PROT-) PROTEGENE INC.
 PA (SAGA) SAGAMI CHEM RES CENT.
 XX Kato S, Sekine S;
 PI Kato S, Sekine S;
 XX WPI: 1999-264019/22.
 DR N-PSDB; AAX34737.
 XX Human proteins with secretory signal sequences and nucleotide
 PT sequences, useful in control of proliferation and differentiation of
 PT cells

XX

PS This DNA encodes a protein having homology to human complement factor H,
 CC which plays a role in the immunological mechanism involving the
 CC complement reaction. The protein can also be used as an antigen for
 CC preparing antibodies against the protein. The cDNA can be used as a probe
 CC for gene diagnosis and the gene for gene therapy, as well as for large-
 CC scale expression of the protein. The protein may also have immune
 CC stimulating or suppressing activity, haematopoiesis regulating activity,
 CC tissue growth activity, activin/inhibin activity, anti-inflammatory
 CC activity, tumour inhibition activity, chemotactic/chemokinetic activity,
 CC receptor/ligand activity, etc. The protein is identified by the presence
 CC of a hydrophobic N-terminal secretory signal region, knowledge of the
 CC protein function is not required, as in e.g. methods of expression
 CC cloning.
 XX
 SQ Sequence 578 AA;
 Query Match 31.5%; Score 1001.5; DB 20; Length 578;
 Best Local Similarity 36.7%; Pred. No. 7.1e-72;
 Matches 219; Conservative 94; Mismatches 238; Indels 45; Gaps 19;
 QY 1 MLLSFVILLISWSTVGG-EGTLCDDPKIHGGLYDEEDYNPFSQVPTGTVFYSEYNF 59
 Db 1 MLLINVILTLYWYSCANGQEVKPCDFPEIQHGLYKSLRRLYFPAAAGQSYSYCDQNF 60
 QY 60 VPSKSFWRITCTEAGSPKPKCLRMCSFFV-KNGHSESSGLIHLEGGDTVOICNTG 117
 Db 61 VTFSGSYWDYIHTQDGSPTVCLATCKSDVEIEINGFISESSYIILNEETQYNCKPG 120
 QY 118 YSLQ--NNEKNISCVBERGWSPTTSCFTKGECHVPILEAVDAQPKKESYKVGDLKFS 174
 Db 121 YATADGNSSGITCLQNGWSTQPIKFIK-----CDMPVE-NSRAKSNMFKLHDLTDYE 175
 QY 175 C---RKNLIRVSDSVQCYOFQWSPNFPKCKQVRSCGPPPLNSGEVKEIRKEEYGHNE 231
 Db 176 CYDGYESSYGNTTDSIVCGEDGWS-HLPTCYNSESCGPPPTISNGDTTSFPOKVLPS 234
 QY 232 VREYDCNPNFIINGPKKICVDGEMTTLPTCVBQVKTGCVIPELYG--YVQPSVPPY-- 287
 Db 235 RVEYQCYSEYLOGSKYVTCNSGDNSEPPRCI-SMKPCEP-PEIQHGLYENTRRPYFP 292
 QY 288 -QHGVSVEVNCREYAMIGN---NMITCINGIWTPLMPCVATHQKRCCKIAGVNIKT-LL 342
 Db 293 VATGQSYSYVCDQNFVTPSGSYWDYIHTQDGLPTVPC-----LRTCKSDIEIENGFI 347
 QY 343 KLSGKEFNHNSRIYRCSDFRYRHS-----VCINGKNPEVDCTEKREQFCPPPPQIP 396
 Db 348 SESSSYIILNKETQYKCKPGYATADGNSSGITCLQNGWSAQPICTI----KFC-DMPVFE 402
 QY 397 NAQNMTTTVNYQDGEKVAVLCKENYLLPEAK---EIVCKDGRWQSLPRVSTAYCGPPP 453
 Db 403 NSRAKSNMFKLHDLTDYECDYGEISYGNVTGIVCGEDGWSHPTCTYNSEKCGPPP 462
 QY 454 SINNGDTTSFPLSVYPPGTVTRCQSYFKLOGSVTVTCRNKQWSEPPRCCLDPCVYVSEEN 513
 Db 463 PISNGDTTSFLLKLVVVPQSRVEYQCYSEYVQVQSYVTCNSGSEWSEPPRCIHPICITEEN 522
 QY 514 MNKNNIOLKWRNDGKLYAKTGDAVEFOCKFPHKAMISSPPFRAICQEGKEEYPTICE 569
 Db 523 MNKNNIOLKGRKSDIKYIYAKTGDTIEFMCKLGYNANTSVLSFOAVCREGIVEYPRCE 578

RESULT 8

AA082921
 ID AAP82921 standard; protein; 660 AA.
 AC AAP82921;
 XX AAP82921;
 DT 26-NOV-1990 (first entry)
 DE B subunit of human Factor XIII.

XX	B sub-unit; factor XIII; blood clot stabilisation; fibrin polymer;
KW	crosslink; scleroderma; haemorrhage; ulcerative colitis.
XX	Homo sapiens.
OS	
XX	
FH	Key Location/Qualifiers
FT	Peptide 1..19
FT	/label=leader_sequence
FT	Peptide 20..660
FT	/label-mature_b_subunit
ET	
XX	
PN	AU8778694-A.
PX	
XX	
PD	31-MAR-1988.
XX	
PF	21-SEP-1987; 87AU-0078694.
XX	
PR	19-SEP-1986; 86US-0909512.
XX	
PA	{Zymo-} ZYMOGENETICA INC.
PI	Davie E, Seale RL, Ichinose A, Holly JA, Parker GE;
DR	WPI; 1988-140637/21.
DR	N-PSDB; AAN84202.
XX	
PT	New DNA sequences encoding sub-units of factor 13 - and corresponding
PT	expression vectors and transformed host cells.
XX	
PS	Disclosure; ; P: English.
XX	
CC	The b subunit starts with an amino-terminal sequence of Glu-Glu-Lys-Pro,
CC	earlier reported by Takagi and Doolittle, Biochem 13:750-756, 1974.
CC	The partial leader sequence includes a typical hydrophobic core and
CC	an Ala residue at position 19 and Leu at position 17.
CC	The protein sequence predicted from the cDNA includes two potential
CC	Asn-linked glycosylation sites with a sequence of Asn-Tyr-Ser or
CC	Asn-Gly-Ser, starting with amino acid residues 161 and 544 resp.
CC	In addition, a third potential carbohydrate attachment site is
CC	present in the sequence of Asn-Arg-Cys, starting at residue 271.
CC	The differential glycosylation of those Asn residues may in part
CC	be responsible for the microheterogeneity of the a subunit.
CC	There is considerable internal gene duplication, involving ten
CC	repetitive sequences of ca. 60 amino acids, they are very similar to
CC	three repeated segments present in the Ba chain of factor B of human
CC	complement.
CC	When cultured the host cells will produce the polypeptide which
CC	can be assembled to factor 13, which stabilises blood clots by
CC	crosslinking fibrin polymers. Factor 13 is useful therapeutically, e.g.
CC	in cases of scleroderma, haemorrhage, ulcerative colitis etc., and
CC	can be prepd. in large amts. without risk of viral contamination.
CC	See also AAN82401 and AAN82403.
XX	
SQ	Sequence 660 AA;
	Query Match 23.8%; Score 756.5; DB 9; Length 660;
	Best Local Similarity 28.8%; Pred. No. 4.3e-52;
	Matches 194; Conservative 85; Mismatches 255; Indels 139; Gaps 21
OY	3 LLEFVILISWSTVGEGTCLDFPKIHGHFLYDEEDYNPFSQVPTGEV--EYYS----- 54 : : : : : : : : :
Db	5 LTFIILI-----ISGE-----LYAEKPCGGPPHVENGRIAQYYITFKSYF 46 : : : : : : : : :
OY	55 -----CEYNFVSPKSFWTRITCTEEGWSPTPKLCRMCSFPVKNGHESSGLI 103 : : : : : : : : : : : : : : : : : : : : : : :
Db	47 PMSIDKLSPFCLAGYTSGRQEQTCTEGWSPEPRCPKCKTKDLSLNGYISDVKLL 106 : : : : : : : : : : : : : : : : : : : : : : :
OY	104 H-----LEG----- 107 : : : : : : : : : : : : : : : : : : : : : : :
Db	107 YKIQENMHYGCASGYKTGTGKDREVOCLSDGSWSQPTCRKEHETCLAPELYNGNYSSTQ 166
OY	108 -----DTVOILICNTGYSLONNEK--NISCVGERGWSTTPPICSFTKKGCHVPILLEAVDAQ 159

QY	422	LIPEAKEIVCKDRNQSLPRCVESTAYCGPPPSINNGDTSFPLSVYPPGSGTYYRCQSF	481
Db	179	EMFGDEEVMCLNGNWTEPPQCKDGTGKGGPPPIIDNGDITSFPLSVYAPASSVEYQCQNL	238
QY	482	YKLOGSVTVTCRNKOWSPPRCLDPCCVYVSEENMKNKIQLKWRNDGKLAKT	533
Db	239	YQLEGNKRITCRGOWSEPPKCLIPCVLSREIMYNIARWTAKQLISRT	290
RESULT 11			
AAW39158	ID	AAW39158 standard; Protein; 229 AA.	
XX	AC	AAW39158;	
XX	DT	27-APR-1998 (first entry)	
XX	DE	Clone PRBS3FH2910 #3.1 CFH related protein fragment.	
XX	KW	Complement factor H; tumour associated antigen; renal cancer;	
XX	KW	urogenital cancer; medicament; modulator.	
XX	OS	Synthetic.	
XX	PN	WO9738136-A1.	
XX	PD	16-OCT-1997.	
XX	PF	09-APR-1997; 97WO-0505710.	
XX	PR	06-MAR-1997; 97US-0812481.	
XX	PR	03-APR-1996; 96US-0015083.	
XX	PR	03-APR-1996; 96US-0630048.	
XX	PR	06-MAR-1997; 97US-0038614.	
XX	PA	(BARD-) BARD DIAGNOSTIC SCI INC.	
XX	PI	Enfield DL, Hass GM, Kinders RJ;	
XX	DR	WPI; 1997-512742/47.	
XX	DR	N-PSDB; AAV02794.	
XX	PT	Treating or screening for cancer, e.g. renal or urogenital cancer -	
XX	PT	by modulating or detecting tumour associated human complement Factor	
XX	PT	H related antigen, or nucleic acid encoding it	
XX	PS	Example 6B; Fig 7B; 104pp; English.	
XX	CC	This partial protein is found in clone PRBS3FH2910 #3.1 and represents a	
XX	CC	complement factor H related protein with homology to a region of the	
XX	CC	human tumour-associated complement factor H (CFH). The detection of this	
XX	CC	protein and a CFH antigen can be used in screening or for the treatment	
XX	CC	of renal or urogenital cancer, e.g. bladder, cervical or prostate cancer.	
XX	CC	Agents that may modulate this antigen could be used in the manufacture of	
XX	CC	a medicament for the treatment of a tumour cell.	
XX	SQ	Sequence 229 AA;	
Query Match 14.6%; Score 456; DB 18; Length 229;			
Best Local Similarity 32.3%; Pred. No. 2.3e-29;			
Matches 96; Conservative 42; Mismatches 83; Indels 76; Gaps 6			
QY	225	EEVGHNEVVEYDCNPFILINGPKKIQCVDGEMTTLPTCBEQVKTGVIPELEYGYQPSV	284
Db	2	DSYQGEETVYKCFEGIDGPAIAKLGKWSHPPSCKIK-TDCPSLPSFE-----NAI	54
QY	285	PP-----YQHGVSVSEVNCREYAMIGNMIMTCINGIWTLPWCVATHQLKRCCKIAGVNI	338
Db	55	PWGEKKVLYKAGEQYTYTCATYCYQMDGASVNTCLNSRWTGRPTC-----	98
QY	339	KTYLLLSGKFEFNHNSRIYRCSDFIRYRHSVNCINWKNPEVDCTEKREQFCPPPPQIPNA	398

Db 75 KYPSEVRYQCSYPMFDEEVMCLNGNWTPEPQCKDSTGKCGPPPPIDNGDITSPFL 134
Qy 466 SVYPPGTVYRQSFYKLOGSVYTCRNKNOWSEPPRCCLDPCVVSSENNKNNITQKW 523
Db 135 SVYAPASSVEYQCONLYQLEGKRNKRTCRNQWSEPPKCLHPCVLSREIMENYIALRW 192

RESULT 13
AAR11139
ID AAR11139 standard; Protein; 1087 AA.

XX AC AAR11139;
XX DT 23-MAY-1991 (first entry)
XX DE B lymphocyte membrane glycoprotein CR2.
XX KW CR2; B lymphocyte; membrane receptor protein; Epstein-Barr virus;
XX CC extracellular domain.
XX OS Homo sapiens.

XX FH Key Location/Qualifiers
FT Region 20..90
FT /note= "1"
FT Region 20..153
FT /note= "2"
FT Region 20..276
FT /note= "3"
FT Region 20..652
FT /note= "4"
FT Region 20..776
FT /note= "5"
FT Region 20..1025
FT /note= "6"
FT Peptide 1..20
FT Protein 21..1087
FT /label= sig_peptide
FT /label= mat_protein

XX W09102536-A.
XX 07-MAR-1991.
XX 22-AUG-1990; 90W0-US04817.
XX 23-AUG-1989; 89US-0398224.
XX (SCRI-) SCRIPPS CLINIC & RE.

XX Moore MD, Cooper ND, Nemerow GR;
XX WPI; 1991-087109/12.
XX N-PSDB; AAQ10989.

XX Synthetic polypeptide(s) of extracellular domain - treat
XX Epstein-Barr virus infections and diagnose same by formation of
XX complex between CR2 and EBV

XX Disclosure; Fig 9.1-9.9; 60pp; English.

XX The six indicated fragments encode claimed and disclosed
XX polypeptides, which are synthesised by recombinant expression, pref.
XX in a baculovirus expression system in which a DNA plasmid, contg. a
XX cDNA encoding this sequence, is truncated to encode a CR2 polypeptide
XX that comprises a region of the extracellular domain of CR2. The
XX construct is then inserted downstream of an appropriate promoter in
XX a transfer vector and integrated into a baculovirus which is then
XX used to infect host insect cells for expression.
XX The polypeptides correspond to B-lymphocyte membrane receptor
XX protein for Epstein-Barr Virus (EBV) and give specific binding.
XX Protein number 3 is used to inhibit infection of mammalian cells in

CC contact with an aq. medium, (esp. mammalian blood) and of human
CC B-lymphocytes, by EBV. The peptides are used to detect the presence
CC of EBV in an aq. sample and to detect antibodies directed against CR2.
CC The peptides are used in a pharmacological compsn. as active
CC ingredient with a carrier to treat immune disorders.
XX
SQ Sequence 1087 AA;

Query Match 12.8%; Score 408; DB 12; Length 1087;
Best Local Similarity 21.9%; Pred No. 8.7e-24;
Matches 170; Conservative 87; Mismatches 246; Indels 272; Gaps 38;

Qy 26 PKIHGFLYDEEDYNFSPQVPTGEVFFYSCYFNVSFWSKFWTRITCTEFG-WSPTP--- 81
Db 158 PMIHNGHHTSEN---VGSIAFGLSVTYSCEGYLLVGEKI---INCLSSGKKSAPPTC 210
Qy 82 ---KCLRMCSPFPVKNHSESSGLIHLEGTQVQIICNTGYSLONNEKNISCVBERG---W 134
Db 211 EEARCKSLGRFP---NGKVKEPPIIRV-GVTANFFCDEGYRLQGPSS-RCVIAGOGVAV 265
Qy 135 STPPICSTKGECHVPIL-ANV-----DAOP----- 160
Db 266 TKMPVCEBIFCPSPPIILNGRHIGNSLANVSYGSIYTYCDPDPBEGVNEILIGESTLRC 325
Qy 161 -----KKEYKVDGVKFSCKRNLIIR 181
Db 326 TVDSQKTGWSGAPRCELSTSAVQCPHPQILRGRVMSGQKDRYTYNDTVIPACMFGLT 385
Qy 182 VGSDSVQCYQFG-WSFNFTCKQVRSQGPPLSGEVEKEIRKEEYGHNEVVEYDCNPN 240
Db 386 KGSQKQIRCAOQGTWEPAPVCE---KECQAPPNILNGQKEDRHMRVDFDGTSTIKYSCNPG 442
Qy 241 FIINGPKIQCV-DGEWT-TLPTCQVQVTC---GYIPELYGVVQVSV----- 284
Db 443 YVLVGEESIQCTSEGVTTPVPOC---KVAACEATGRQLLTTPKPOHOFVRDVSNCSEGYK 500
Qy 285 -----PP-----YQHGVSVEVNCRN-- 299
Db 501 LSGSVYQECQGTIPWFMELRCKEITCPPPPVIYNGAHTGSSLEDPPYGTVTYTCNPGP 560
Qy 300 ---EYAMIGNMITCIN-----GIWT-ELPMCVATHQLKCKIAGVNIKTLKLSKE- 348
Db 561 ERGVESLIGESTIRCTSNQDQERTGWSGAPAPCKLSLLAVQC---SHVHIANGYKISGKEA 618
Qy 349 -FNHNSRIRYRGSDIFRYRHSVCINGK---MNEVDCTEKRQFCPPPPQIPNAQNM-T 402
Db 619 PYFYNDTVTFKYSVGTGKSSQIRCKRDNTPDPEIPVCEKG---CQPPGLHHGRTGG 675
Qy 403 TTVNYODGKAVLCKENYLLPEAKEIVC-KGRWO-SLPRCVES----- 445
Db 676 NTVFFVSGMTVDYTCDPGYLLVGNKSIHCMPSGNWSPSAPRCBETCQHVRSQSLQELPAGS 735
Qy 446 -----TAYCGPPPSINNGDTTSFPLS 466
Db 736 RVELVNTSCODGYQLTGHAQYMCQDAENGIMFKKIPLCVKVICHPPPPVINGKHTGMABE 795
Qy 467 VYPPGSTVYTRC-QSFYKL-QGSVTVTCRNKOW---SEPPRCL-----DPCVVSENNMK 516
Db 796 NELYNGEVSVECDQGYLLGKNCSEVILKAWILERAFOCLRLSLCPNPEVKGHYKLNK 855
Qy 517 -----NNIQLKWRNDGKLYAKTGDAVEFOCKFPH-----KAMISSPP 553
Db 856 THSAYSHNDIVVDCNPG--FIMNGSRV-IRCHTDNTWPGVPTCIKKAFIGCPP 907

RESULT 14
AAG66398
ID AAG66398 standard; Protein; 979 AA.
XX
XX AAG66398;
XX AC
XX 16-OCT-2001 (first entry)
XX

DE Receptor 222.
KW Receptor 222; cytostatic; anti-HIV; cancer; HIV infection.
XX Unidentified.
OS
XX CN1296952-A.
PN
XX 30-MAY-2001.
XX
XX 23-NOV-1999; 99CN-0124086.
XX
PR 23-NOV-1999; 99CN-0124086.
XX
XX (SHAN-) SHANGHAI BORONG GENE DEV CO LTD.
PA
PI Mao Y, Xie Y;
XX
XX WPI; 2001-483894/53.
DR N-PSDB; AAH75787.
XX
XX New polypeptide-complement receptor 222 for treating diseases, such as,
PT cancer and human immunodeficiency virus infection -
XX
XX Claim 1; Pages 21-23 (Disclosure); 30pp; Chinese.
XX
XX The present sequence is the protein sequence for receptor 222. Receptor
CC 222 and its coding sequence are useful in treating diseases e.g. cancer
CC and HIV infection.
XX
XX Sequence 979 AA;
SQ
Query Match 12.4%; Score 395; DB 22; Length 979;
Best Local Similarity 23.2%; Pred. No. 8.4e-23;
Matches 159; Conservative 92; Mismatches 242; Indels 192; Gaps 39;
QY 23 CDFPKI-HHGFLYDEEDYNFSPQVPTGEVFFYSCYFNVPFSKSFTRITCTEAGS-PT 80
DB 125 CDLPAPENGFL-----RTETSMGSAVOYSCPKPGHILAGSDL--RLCLENRKWSGAS 175
QY 81 PKCLRM-CSFPF-VKNGHSESSGLIHLEGDVVOITCNTGYSLONNEKNISCVVERGW-STP 137
DB 176 PRCEAISCKKPNPVMNGSIKSNLYL--STLYECDPGYVLNGTERRTCTQDDKNWDEDE 233
QY 138 PICSTKGECHVPPILEANVDAQPKKESYKVGDLKFSCKRNLIIRVGSVQCYQFG-WSP 196
DB 234 PIC--IPVDCSSPPYSAN--QGVGDEYTFQKEIYETCNEGFLLEGARSRVCLANGWSG 289
QY 197 NFPTCKGQVRSCGPPQLNSGEVKEIRKEEYGHNEVVEYDCNPNFIINGPKKIQC-VDGE 255
DB 290 ATPDCV-PVR-CATPPQLANGVTEGL---DYGFMKEVTFHCHEGYILHGAPKLTQSDGN 344
QY 256 W-TTLPTCEQVKTGCIYIPELEYGVQVSPYQHGVSFVNCREYAMIGNNMTICI-N 313
DB 345 WDAEIPLC--KPVNCGPPEDLAHGF--PNGSFHGHIOYQCFPGYKHLGNSRRCLSN 400
QY 314 GIWT-ELPMCVATHQKRCIKAGVNIKTLKLSGKEFNHNSRIYRCSDFRY----- 365
DB 401 GSWSSGSSCLPC-----RCSPTVIEYGI--VNGDFDQGAARQIOCFKGLKLLSEIT 453
QY 366 -----RHSYC-----INGKW-----NPEVDCT 382
DB 454 CEADQWSSGPPHCEHTSGSLPMIPNFISETSSKENVITYSCRSYVIOGSSDLICT 513
QY 393 EK-----REQFCPPPIQIPNA----- 398
DB 514 EKGWMSQYPVCEPLSCGSPSVANAVATGEAHTYSEVKLRCLBEGYTMDDTDTFTCOK 573
QY 399 -----QNMFTTVNYOD---GEKVAVLCKENYLLPEAKEIYCK-DG 434
DB 574 DGRWTFPERISCPKCPIDENITHILVHGDDDFSVNRQVSVSACGEGTTEGVNISVCQLDG 633
QY 435 RMQ-----SLPRCVESTAYCGPPPSINNGDTSFPLSVYPPGSTVTYRCSFYKQLQGSVTVT 491

DB 634 TWEPFSDSC--SPVSCGKPESPHGFVWG---SKYTFESTIIYQCEPGYELEGNRERV 688
QY 492 CR-NKOWS-----EPPRCLDPCVVSEENMN-KNNIQLKWRNDGKLYAKTGDAVEFQCKF 543
DB 689 COENRQWGGVAICKETRCETPL-----EFLNGKADIENR-----TTGPNVVYSC-- 733
QY 544 PHKAMISPPFERAICQE-GKFEYPI 567
DB 734 -NRGYSLEGPSEAHCTENGTSHPV 757
RESULT 15
ABP37987
ID ABP37987 standard; Protein; 1066 AA.
XX
AC ABP37987;
XX
XX 23-JUL-2002 (first entry)
XX
XX Human GS96663 protein.
DE
XX Human; chromosome 9q31-34; lipoprotein metabolism disorder;
KW cholesterol transport disorder.
XX
XX Homo sapiens.
OS
XX WO200071710-A2.
PN
XX 30-NOV-2000.
PD
XX 25-MAY-2000; 2000WO-FR01426.
PF
XX 25-MAY-1999; 99FR-0006587.
PR 16-JUN-1999; 99US-0139450.
XX
XX (AVET) AVENTIS PHARMA SA.
PA
PI Denefle P, Rosier-Montus M, Arnould-Reguigne I, Prades C, Clepet C;
XX WPI; 2001-025161/03.
DR N-PSDB; ABN93421.
XX
XX New nucleic acid derived from human chromosome 9, used e.g. for
PT diagnosis and drug screening, derived from genes implicated in
PT disorders of lipoprotein metabolism -
XX
XX Claim 22; Page 262-266; 269pp; French.
PS
XX The present sequence is the protein sequence encoded by a human gene from
CC chromosome 9q31-34. This sequence is likely to be involved in diseases
CC of plasmatic lipoprotein metabolism, e.g. the reverse transport of
CC cholesterol.
XX
SQ Sequence 1066 AA;
Query Match 12.4%; Score 395; DB 22; Length 1066;
Best Local Similarity 23.2%; Pred. No. 9.4e-23;
Matches 159; Conservative 92; Mismatches 242; Indels 192; Gaps 39;
QY 23 CDFPKI-HHGFLYDEEDYNFSPQVPTGEVFFYSCYFNVPFSKSFTRITCTEAGS-PT 80
DB 212 CDLPAPENGFL-----RTETSMGSAVOYSCPKPGHILAGSDL--RLCLENRKWSGAS 262
QY 81 PKCLRM-CSFPF-VKNGHSESSGLIHLEGDVVOITCNTGYSLONNEKNISCVVERGW-STP 137
DB 263 PRCEAISCKKPNPVMNGSIKSNLYL--STLYECDPGYVLNGTERRTCTQDDKNWDEDE 320
QY 138 PICSTKGECHVPPILEANVDAQPKKESYKVGDLKFSCKRNLIIRVGSVQCYQFG-WSP 196
DB 321 PIC--IPVDCSSPPYSAN--QGVGDEYTFQKEIYETCNEGFLLEGARSRVCLANGWSG 376
QY 197 NFPTCKGQVRSCGPPQLNSGEVKEIRKEEYGHNEVVEYDCNPNFIINGPKKIQC-VDGE 255

Db 377 ATPDCV-PVR-CATPQLANGVTEGL---DYGFMEKVTFFHCHGYILLHGAPKLTQCSQDGN 431
Qy 256 W-TTLPTCVEQVKTGYPYQHVQPSVPPYQHVSVVNCNREYAMIGNNMITCI-N 313
Db 432 WDAEIPLC--KPVNCGPPEDLNGF--PNGFSFIHGHHIQYQCFPGYKLGHSRRCLSN 487
Qy 314 GIWT-ELPMCVAHQLRKCIAGVNIKTLLKSGKEFNHNSRIRYRCSIDIFRY----- 365
Db 488 GSWGSSPSCLPC---RCSTPVIEYCT---VNGTDFDGCKAARIQCFKGLLGLSEIT 540
Qy 366 -----RHSVC-----INGKW-----NPEVDCT 382
Db 541 CEADGQWSSFPHCEHTSCGSLPMIPNAFISSETSSWKENVITYSCRSYVIQSSDLICT 600
Qy 383 EK-----REQFCPPPPQIPNA----- 398
Db 601 EKGWWSQPYVCEPLSCGSPSPVANAVATGEAHTYSEVKRLCLEGYTMDTDTDTFTCQK 660
Qy 399 -----QNMTTVNYQD-----GEKVAVLCKENYLLPEAKEIVCK-DG 434
Db 661 DGRWFPERISCPKCKPLPENITHILVHGDDFSVNRQVSVSCAEGYTFEGVNISVCOLDG 720
Qy 435 RWQ---SLPRCVSTAYCGPPPPINNGDITSPPLSVYPPGSTVYRCQSFYKLGQSVTVT 491
Db 721 TWEPPFSDSC---SPVSCGPSPGPHGVVG---SKYTFESTIYQCEPGYELEGNRERY 775
Qy 492 CR-NKQWS-----EPPRCCLDPCVSENNM-KNNIQLKWRNDGKLYAKTGDAVEFOCKF 543
Db 776 QOENROWSGGAICKETCETPL-----EFLNGKADIENR-----TTGPNVWVYSC-- 820
Qy 544 PHRAMISSPPFRAICQF-GKFEYPI 567
Db 821 -NRGYSLEGPSEAHCTENGTSHPV 844

Search completed: July 17, 2003, 13:50:54
Job time : 77 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2003, 13:49:33 ; Search time 85 seconds
(without alignments)
1379.304 Million cell updates/sec

Title: US-09-805-337A-2

Perfect score: 3183

Sequence: 1 MLLFSVILISWSTVGEG.....SPPFRAICQGRFEYPICE 569

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21.*

1: sp_archaea.*

2: sp_bacteria.*

3: sp_fungi.*

4: sp_human.*

5: sp_invertebrate.*

6: sp_mammal.*

7: sp_mhc.*

8: sp_organelle.*

9: sp_phage.*

10: sp_plant.*

11: sp_rodent.*

12: sp_virus.*

13: sp_vertebrate.*

14: sp_unclassified.*

15: sp_rvirus.*

16: sp_bacteriap.*

17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3183	100.0	569	4	Q9BXR6	Q9bXR6 homo sapien
2	1524	47.9	808	11	Q61408	Q61408 mus musculus
3	1463.5	46.0	1236	11	Q91YB6	Q91YB6 rattus norv
4	1426	44.8	699	11	Q91WX0	Q91WX0 rattus norv
5	1355.5	42.6	1172	4	Q9NU87	Q9nu87 homo sapien
6	1284	40.3	657	4	Q14006	Q14006 homo sapien
7	906.5	28.5	343	11	Q61406	Q61406 mus musculus
8	693.5	20.8	509	11	Q8R018	Q8r018 mus musculus
9	646	20.3	452	11	Q61407	Q61407 mus musculus
10	476	15.0	1053	13	Q91275	Q91275 paralabrax
11	435.5	13.7	1045	6	Q46545	Q46545 ovis aries
12	393.5	12.4	3567	11	Q9ES77	Q9es77 mus musculus
13	387.5	12.2	303	11	Q61405	Q61405 mus musculus
14	387.5	12.2	1911	6	Q29528	Q29528 papio hamad
15	387.5	12.2	2014	6	Q29530	Q29530. pan troglod
16	383	12.0	2039	4	Q16745	Q16745 homo sapien

17	383	12.0	2489	4	Q16744	Q16744 homo sapien
18	382	12.0	679	11	Q99254	Q99254 mus musculus
19	382	12.0	1032	11	Q9DC83	Q9dc83 mus musculus
20	380.5	12.0	3389	4	Q96Q09	Q96qu9 homo sapien
21	379.5	11.9	645	12	Q9WRU2	Q9wru2 macaca mula
22	362	11.4	3508	4	Q96RM4	Q96rm4 homo sapien
23	355	11.2	661	6	Q29531	Q29531 pan troglod
24	353	11.1	522	6	Q28769	Q28769 papio cynoc
25	352	11.1	195	4	Q8WW96	Q8ww96 homo sapien
26	342	10.7	1124	5	Q9VVR4	Q9vvr4 drosophila
27	341	10.7	946	4	Q96Q03	Q96q03 homo sapien
28	340.5	10.7	579	11	Q60736	Q60736 mus musculus
29	338	10.6	3564	11	Q923L3	Q923l3 mus musculus
30	336	10.6	974	5	P91658	P91658 drosophila
31	333	10.5	522	6	Q28797	Q28797 pan troglod
32	332.5	10.4	669	11	Q922H0	Q922h0 mus musculus
33	330	10.4	555	11	Q99JA1	Q99ja1 cavia porce
34	324.5	10.2	601	11	Q9CUT3	Q9cut3 mus musculus
35	323.5	10.2	559	4	Q9UQV2	Q9uqv2 homo sapien
36	322.5	10.1	559	11	Q63135	Q63135 rattus norv
37	311	9.8	533	11	O08569	O08569 cavia porce
38	300.5	9.4	469	11	Q91X48	Q91x48 mus musculus
39	294.5	9.3	417	11	O35520	O35520 rattus norv
40	294.5	9.3	1853	5	Q9VIU9	Q9viu9 drosophila
41	281	8.8	1875	5	Q93691	Q93691 caenorhabdi
42	280.5	8.8	481	4	Q9H284	Q9h284 homo sapien
43	276.5	8.7	497	11	Q63612	Q63612 rattus norv
44	268	8.4	610	6	Q95LG1	Q95lgl equus cabal
45	266.5	8.4	2043	5	Q96943	Q96943 geodia cydo

ALIGNMENTS

RESULT 1

Q9BXR6	PRELIMINARY;	PRT;	569 AA.
ID	Q9BXR6		
AC	Q9BXR6;		
DT	01-JUN-2001 (TREMBLrel. 17, Created)		
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)		
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)		
DE	Complement factor H-related protein 5.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCHI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=21125890; PubMed=11058592;		
RA	McRae J.L., Cowan P.J., Power D.A., Mitchellhill K.I., Kemp B.E.,		
RA	Morgan B.P., Murphy B.F.;		
RT	"Human Factor H-related Protein 5 (FHR-5). A New Complement-Associated		
RT	Protein.";		
RL	J. Biol. Chem. 276:6747-6754(2001).		
DR	EMBL; AF295327; AAK15619.1; -.		
DR	HSSP; P10998; 1VVD.		
DR	InterPro: IPR000436; Sushi_SCR_CCP.		
DR	Pfam; PF00084; sushi; 8.		
DR	SMART; SM00032; CCP; 8.		
SQ	SEQUENCE 569 AA; 64419 MW; 7FAAE31707B0C112 CRC64;		

Query Match 100.0%; Score 3183; DB 4; Length 569;
Best Local Similarity 100.0%; Pred. No. 9.5e-275; Indels 0; Gaps 0;
Matches 569; Conservative 0; Mismatches 0;

Qy 1 MLLFSVILISWSTVGEGTLCDFPKIHGHGLYDEEDYNPFQVPTGEVYSCYXNFV 60

|||||

1 MLLFSVILISWSTVGEGTLCDFPKIHGHGLYDEEDYNPFQVPTGEVYSCYXNFV 60

|||||

61 SPKSFVTRITCTEGWSPTPKLRMCSFPVKNHSESSGLIHLEGTVOICNTGYSL 120

|||||

61 SPKSFVTRITCTEGWSPTPKLRMCSFPVKNHSESSGLIHLEGTVOICNTGYSL 120

```

QY 121 ONNEKNISCVGRHSTPPICSFTKGECHVPILEANDVDAQPKKESYKVGDLVKFSCRKNLI 180
Db 121 ONNEKNISCVGRHSTPPICSFTKGECHVPILEANDVDAQPKKESYKVGDLVKFSCRKNLI 180
QY 181 RVGSDSVQCYQFGWSPNFTCKGQVRSCTGPPOLSGEVEKIRKEEYGHNEVEYDCNPN 240
Db 181 RVGSDSVQCYQFGWSPNFTCKGQVRSCTGPPOLSGEVEKIRKEEYGHNEVEYDCNPN 240
QY 241 FIINGPKKIQCVGEMTTLPTCQEVQKTCGYIPELEYGYVQSPVPPYOHGVSVEVNCNE 300
Db 241 FIINGPKKIQCVGEMTTLPTCQEVQKTCGYIPELEYGYVQSPVPPYOHGVSVEVNCNE 300
QY 301 YAMIGNNMITCINGIWTLPWCVATHOLKCKTAGVNIKTLLKLSGKEFNHNSRIYRCS 360
Db 301 YAMIGNNMITCINGIWTLPWCVATHOLKCKTAGVNIKTLLKLSGKEFNHNSRIYRCS 360
QY 361 FIINGPKKIQCVGEMTTLPTCQEVQKTCGYIPELEYGYVQSPVPPYOHGVSVEVNCNE 420
Db 361 FIINGPKKIQCVGEMTTLPTCQEVQKTCGYIPELEYGYVQSPVPPYOHGVSVEVNCNE 420
QY 421 YLLPEAKEIKVCKDRMQSLPRCVSTAYCGPPPSINNGDTSFPLSYPPGSTVYRCQS 480
Db 421 YLLPEAKEIKVCKDRMQSLPRCVSTAYCGPPPSINNGDTSFPLSYPPGSTVYRCQS 480
QY 481 FYKLGQSVTTCRNKOWSEPPRCCLDPCVYSEENMKNNTQKWRNDGKLYAKTGDAVEFQ 540
Db 481 FYKLGQSVTTCRNKOWSEPPRCCLDPCVYSEENMKNNTQKWRNDGKLYAKTGDAVEFQ 540
QY 541 CKFPHKAMISSPPFRAICQEGKFEYPICE 569
Db 541 CKFPHKAMISSPPFRAICQEGKFEYPICE 569

RESULT 2
ID Q61408 PRELIMINARY; PRT; 808 AA.
AC Q61408;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Complement factor H-related protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90153969; PubMed=1689298;
RA Vik D.P., Munoz-Canoves P., Kozono H., Martin L.G., Tack B.F.,
RA Chaplin D.D.;
RT "Identification and sequence analysis of 4 complement factor H-related
RT transcripts in mouse liver.";
RL J. Biol. Chem. 265:3193-3201(1990).
DR EMBL; M29009; AAA37416.1; -
DR HSSP; P08603; 1HCC.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 13.
DR SMART; SM00032; CCP; 13.
SQ SEQUENCE 808 AA; 91654 MW; 6FD97D53CE74DF6D CRC64;

Query Match 47.9%; Score 1524; DB 11; Length 808;
Best Local Similarity 36.2%; Pred. No. 6.4e-127;
Matches 292; Conservative 87; Mismatches 187; Indels 240; Gaps 6;

QY 1 MLLFSVLISWSTVGGEGTLCDFPKIHGHFLYDEEDYNPFESQVPTGVEFVYSCYENFV 60
Db 6 MLLSNILLTAMLSAKGEVKSCFEPQFKYGRLYEEILRPFPVYSIGNKYKCDNGFSP 65
QY 61 SPSKSFWRITTEGWSFTPKCLRMCSFPFKVKNHSESSGIIHLEGDTVQICNTGYSL 120
Db 66 PPSGLFWLDYIRCTVGVGKPEVCRKCVFHYVVEGEFAFWKXIYVQGSQSLKVCYNGYSL 125
QY 121 QNNEKNISCVGRHSTPPICTC-----SF----- 142

```

```

Db 126 QNGQDTMTCTENGWSPPPKIRIKTCSVSDIEIENGFFSEFRTYALNRETYSRCKQYV 185
QY 143 ----- 142
Db 186 TMTGTSRSITCLONGWSPPOSCIKSCERPVSNTKNNSTWFKLNDKLDYECLIGHEN 245
QY 143 -----TKG-----ECHVPFILEANDVDAQPKKESYKVGDLVKFSCRKNLI 180
Db 246 EYKHTKGSITCTYGVGSDTSCPYEIECSVPIIDRLKLVSPRKEKYRGVLDLLESCRSR-H 304
QY 181 RVGSDSVQCYQFGWSPNFTCKGQVRSCTGPPOLSGEVEKIRKEEYGHNEVEYDCNPN 240
Db 305 RVGSDSVQCYQFGWSPNFTCKGQVRSCTGPPOLSGEVEKIRKEEYGHNEVEYDCNPN 364
QY 241 FIINGPKKIQCVGEMTTLPTCQEVQKTCGYIPELEYGYVQSPVPPYOHGVSVEVNCNE 300
Db 365 FLKLGPNKIQCVGEMTTLPTCQEVQKTCGYIPELEYGYVQSPVPPYOHGVSVEVNCNE 424
QY 301 YAMIGNNMITCINGIWTLPWCVATHOLKCKTAGVNIKTLLKLSGKEFNHNSRIYRCS 360
Db 425 FTMIGLSVSLSGKWTQLPKCVATDLEKCRVLSKSTDIETAIKPRNEFOHNSMTMYKCR 484
QY 361 DIPRYHSHVCINGKWNPEVDCTEKREOFQCPPPPQIPNAQNMNTTVNYQDGEKVAVLCKEN 420
Db 485 DKOEYHSHVCINGKWNPEVDCTEKREOFQCPPPPQIPNAQNMNTTVNYQDGEKVAVLCKEN 542
QY 421 YLLPEAKEIKVCKDRMQSLPRCV----- 443
Db 543 YLTQDPEEMCKDRMQSLPHCLGLPCGPPPSILRGTVLSLESYOHGEEVTHCSTGFG 602
QY 444 ----- 443
Db 603 IDGPAFICEGGKWSDPKCIKTNCVDVLTPIENAILIRGKKKSYRTGEQVTPRCQSPYQM 662
QY 444 -----ESTAYCGPPPSINNGDTSFPLSYPPGSTVYRCQSPFYK 483
Db 663 NGSDTVCVNSRWIGKPVCKDSRGKCGPPPIDNGDITSLSLPEYFPFSSVDYQCKYIL 722
QY 484 LOGSVTTCRNKOWSEPPRCCLDPCVYSEENMKNNTQKWRNDGKLYAKTGDAVEFQCKF 543
Db 723 LKGGKTTTCRNKOWSEPPRCCLDPCVYSEENMKNNTQKWRNDGKLYAKTGDAVEFQCKF 782
QY 544 PHKAMISSPPFRAICQEGKFEYPICE 569
Db 783 GYHKARGSPFRTKCIISGTINYPTCE 808

RESULT 3
QY1YB6
ID Q91YB6 PRELIMINARY; PRT; 1236 AA.
AC Q91YB6;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Complement inhibitory factor H.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=LIVER;
RA Demberg T., Goetze O., Schlaf G.;
RT "Rat complement factor H: molecular cloning, sequencing and expression
RT in tissues and isolated cells.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ320522; CAC67513.1; -
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF00084; sushi; 20.
DR PROSITE; P500028; ZINC_FINGER_C2H2_1; UNKNOWN_1.

```

```
SQ SEQUENCE 1236 AA; 140343 MW; 1AC89FFA28232EBF CRC64;
Query Match 46.0%; Score 1463.5; DB 11; Length 1236;
Best Local Similarity 31.7%; Pred. No. 2.6e-121;
Matches 289; Conservative 84; Mismatches 172; Indels 367; Gaps 4;
Qy 23 CDPKIHGFLYDEEDNPNPQVPTGEVFYISCEYNFVSPKSFVWTRITCTEGWSPTPK 82
Dy 325 CDFPQKHRLYTESRRPYPVPVIGKEYSYCDNGFTTPSQSYWDLRCTVNGWEPEVP 384
Qy 83 CLRMCSPPFVKNHSESSGLIHLEGDVQIICNTGYSLONNEKNISVERGWSPPIC-- 140
Dy 385 CLRCIFHYVEGESLWQRYIEGSAKVCHSGSLPNQDQDILCTENGWSPPKVCR 444
Qy 141 ----- 140
Dy 445 IKTCVSDIEIENGFFSESDYTYALNRKRYRCKQGYVTNTGETSGIITCLQDWSRPS 504
Qy 141 -----SFTKG----- 145
Dy 505 CIKSCDMPVFENAMTKNNWFKLNDKLDYECHEGYENYKHTKGSITCTYDGSWSPSC 564
Qy 146 ---ECHVPILAEANDVADPKKESYKVGDLKFCSCNKLRYGSDSVOCYQFQWSPNFTCK 202
Dy 565 YERCSIPLLHQDLVVPREVYKVGDSLFSRSG--HRVGADLVQCYHFGWSPNFTCE 623
Qy 203 QVRSCGPPQLSNGSEYKEIRKEEYGHNEVEYDCNPNFIINGPKKIQCVDGWTTLPTC 262
Dy 624 QGVKSCDPLEINPGEIKGPKVEYSHGVDVEYDCKPRFLKGNPKLCQVDGKWTTLPIC 683
Qy 263 VEQVKTCGYIPELYGVQSPVPYQHGVSVEVNCNREYAMIGNMIMTICINGIITELPMC 322
Dy 684 VEYERTCGDLPELHSGSVKLSVPPYHHGDSVEFTCTETFTMIGHAVFICISGRWTELPOC 743
Qy 323 VATHOLAKCTAGVNIKTLLKSGEENHNSRIYRCSDFIRYHSHVCINGKNWPEVDC 382
Dy 744 VATDLEKCAKPGSTGIDAHNKNNEHNHNSVSRCKQEQEYHSHICINGRWDPENCT 803
Qy 383 EKRFQCPPPPQIPNAQNTTNYQDGEKAVLCKENYLLPEAKEIVCKDGRWQSLPRC 442
Dy 804 RNEKRFQCPPPPQIPNAQVETTVKYLQGEKSVLCQDGYLTQGPPEVCKHGRWQSLPRC 863
Qy 443 VE----- 444
Dy 864 TEKIPCSOPKIEHGSIKSPRSSEERDLIESSSYEHGTTFSYVCDGDFRISERNRVCNM 923
Qy 445 ----- 444
Dy 924 GKWSLPRCVICPGPPPSIPLGLIVSHELSEYQYGEVYNCSEGFIDGPAFKCVGGQ 983
Qy 445 ----- 444
Dy 984 WSEPPKCIKTDNDLPTFEIAKPTKKKKSYRSGEQVTFRCPPYRMDGSDIVTCVNTKW 1043
Qy 445 ----- 444
Dy 1044 IGQPVCKDNCSVNPVHPNATILTRHKTYPGDKVRYDCMKPPELFWGSGDVPRKGF 1103
Qy 445 -----STAYCGPPPSINNGDTSFPLSVYTPGSTVYRCQSFYKLOGSVTVTCRNKQ 496
Dy 1104 TEPKCKDSTCKCGPPPIDNGDITSLPVPALSSVEYQCNLYLLKGNKIVTCRNKG 1163
Qy 497 WSEPPRCIDPCVSEENNNKNIQLKWRNDGKIYAKTGDAVEFOCKFPFKAMISSPPFRA 556
Dy 1164 WSPQPTCLHACVIPEDIMEXHNIVLRENAKIYSQSGENIEFPMCKPGYRFRKRGSPFRT 1223
Qy 557 ICQEGKFEYPTC 568
Dy 1224 KCIEGHINPTC 1235
RESULT 4
Q91WX0
```

```
ID Q91WX0 PRELIMINARY; PRT; 699 AA.
AC Q91WX0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Complement factor H-related protein.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=KIDNEY;
RA Ren G., Quigg R.J.;
RT "Rat complement factor H-related protein sequence."
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF436847; AAL25802.2;
SQ SEQUENCE 699 AA; 78309 MW; 2CDC1B0D1C1FDA907 CRC64;
Query Match 44.8%; Score 1426; DB 11; Length 699;
Best Local Similarity 44.7%; Pred. No. 2.8e-118;
Matches 263; Conservative 76; Mismatches 174; Indels 75; Gaps 4;
Qy 53 YSEYNFVSPKSFVWTRITCTEGWSPTPKCLRMCSPPFVKNHSESSGLIHLEGDVQI 112
Dy 113 YVCKQGYTTANGERSGSITCLQTGWSAQPSCKSDVPVFENAKSKNDSTWFKLNDKLDY 172
Qy 113 ICNTGY--SLONNEKNISVERGWSPTICSTFGECVHPILAEANDVADPKKESYKVG 170
Dy 173 ECHTGYENYKHTKGSIMCTDGSWSPSC--YEICSIPLDLPRLVVYPRKIYTTGGDL 230
Qy 171 LKFSCKNLIRVGSQVQYQFQWSPNFTCKGQVRSCGPPQLSNGSEYKEIRKEEYCHN 230
Dy 231 LKFSRPG--HRVGSDSVOCYQFQWSPNFTCKGQVRSCGQYPELLNGEIKGTKRAEYSHG 289
Qy 231 EVYEDCNPNFIINGPKKIQCVDGWTTLPTCVQVTCGYIPELEYGVQSPVPYQHG 290
Dy 290 DVVEYDCKPRFLKGNPKIQCVDGKWTTLPTCVBEKRTCGDLPELHSGVSKFSPYHHG 349
Qy 291 VSVVNCNREYAMIGNMIMTICINGIITELPCVATHOLKCKIAGVNIKTLLKSGKEFN 350
Dy 350 DSVEFTCAEFTVLQGSVSCISGRWAQLPRCVATDLEKCAKPLAVIGTNKLLKKGDLN 409
Qy 351 HNSRIRYRCSDFIRYHSHVCINGKNWPEVDCTEKREQFCPPPPQIPNAQNTTNYQD 410
Dy 410 HNSISYKGGSQYGYGICINGRWDPETCTRKEKTCFPPPPQIPNAHVETTVKYL 469
Qy 411 EKVAVLCKENYLLPEAKEIVCKDGRWQSL----- 439
Dy 470 EKSVLCODGYLTQGPPEVCKHGRWQSLPCTAKITCSQPPKIDHGSIKLPLRSDAGRD 529
Qy 440 -----PRCEVSTAYCGPPPSINNGDT 460
Dy 530 AIEYSSHEHGTTFYSVCDGDFRISKENGVTCHMGKWSPPRCVSTGCKGPPPIENGDI 589
Qy 461 TSPPLSVYTPGSTVYRCQSFYKLOGSVTVTCRNKQWSEPPRCIDPCVSEENNNKNI 520
Dy 590 TSLSLPVYAPLSSVEYQCSFYKMGSKKITCKNGEWEPPKCLHACVLTBEIMRRQNI 649
Qy 521 LKWRNDGKIYAKTGDAVEFOCKFPFKAMISSPPFRAICQEGKFEYPTC 568
Dy 650 PKRTKNKIYIOSGDYVEFVCLTAYQKAQGSPEFRTQCIDGHINPTC 697
RESULT 5
Q9NU87
ID Q9NU87 PRELIMINARY; PRT; 1172 AA.
AC Q9NU87;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE DJ177P10.1.1 (H factor 1 (complement) isoform 1).
GN HFL.
```



```

Db 343 WFKLNDKLDYBCLVGFENEXKHTKSGITCTYYGWSDTSCYDSTRTCGPPPIDNGDITS 402
QY 463 FPLSYVPGSTVTVRCOSFYKLOGSVTVTCRNKQWSEPPRCGLDPCVSVSENNKNNIOLK 522
Db 403 LSLPEYELSSVDQOCYYILKGNKTITCRNGKWSEPTCLHACVIPENIMEAHNIILK 462
QY 523 WRNDGKLYAKTGDVAFQCKPFPKAMISSPPFRAICQEGKFEYPC 568
Db 463 WRHTEKIYAHSGEDIEFECKRGYQKARGLSPFRKTCINGTINYPTC 508

RESULT 9
Q61407
ID Q61407 PRELIMINARY; PRT; 452 AA.
AC Q61407;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Complement factor H-related protein (Fragment).
GN A1194696.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Vik D.P., Munoz-Canoves P., Kozono H., Martin L.G., Tack B.F.,
RA Chaplin D.D.;
RL Submitted (APR-1990) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90153969; PubMed=1689298;
RA Vik D.P., Munoz-Canoves P., Kozono H., Martin L.G., Tack B.F.,
RA Chaplin D.D.;
RT "Identification and sequence analysis of 4 complement factor H-related
RT transcripts in mouse liver.";
RL J. Biol. Chem. 265:3193-3201(1990).
DR EMBL; M29010; AAA37415.1;
DR HSSP; P08603; IHFI
DR MGD; MGI:2138169; A1194696.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 7.
DR SMART; SM00032; CCP; 7.
FT NON_TER 452
SQ SEQUENCE 452 AA; 51602 MW; 2B697AF6C6E13CA CRC64;

Query Match 20.3%; Score 646; DB 11; Length 452;
Best Local Similarity 24.9%; Pred. No. 4.3e-49;
Matches 157; Conservative 66; Mismatches 160; Indels 248; Gaps 13;

QY 1 MLLFLSVLLISWYSTVGGEG----- 20
Db 6 MLLLSNLLTAWLSATKGEKCTSPYILNGIYTPHRIIHRSDDEIRYECNYGFVPTGS 65
QY 21 -----TL--CDPPKIHGFLYDEEDYNPFQSVPTGEVFIYSEYFNVSF 62
Db 66 TVSKCTPTGWIPVPRCTLAPCEPPQKRYGLYEEISLRNFPVIGNKYSYRCNDGFSPP 125
QY 63 SKSFWRITCTEEGWSPKRLMCSFPFVKNGHSSGLIHLEGTQVLIICNTGYSLOK 122
Db 126 SGYSMDYLRLCTAQGWEPVPCVKVHYVENGDSAYWEKIVYQGSRLKVCNGYSLOK 185
QY 123 NEKNISCVRGWSTPDCISFTKGECHVPILEANVDAQPKESYKVGDLVLFKSRKMLIRV 182
Db 186 GDIMTCTENGWSPPPK----- 206
QY 183 GSDSVQCYQFGSNPPTCKGVRCGPPPLQSLNGEVKEIRKEEYGHNEVVEYDCNPFI 242
Db 207 -----KTCAS-----DIHIDNGLSP--SSYIYALNRETSYRCQGVY 243
QY 243 INGPK---KIQCVDGWTLPTCTVQVQVTCGYIPELEYGVQSPVPPYGHGVSEVNCRN 299
Db 244 TNTGETSGITCLQNGWSPQSC---IKSC-----DMPVFENSIT----- 280

```

```

QY 300 BYAMIGNMITCINGIWTETELPMCVATHQKCRCKIAGVNIKTLKLSGKEFNHNSRIRYRC 359
Db 281 -----KNTRTWFKL-----NDKLDYEC 297
QY 360 SDIF--BYRHSVCLNGKNWPEVDCTEKREQFCPPPPQIPNAQNMTTIVNVDGKAVILC 417
Db 298 LVGFENEYKHT----- 308
QY 418 KENYLLPEAKEIVCKDGRWQSLPRCVSTAYCGPPPSINNGDTSFPLSVVPPGSTVYR 477
Db 309 -----KGSITCTYYGWSDTSCYDSTRTCGPPPIDNGDITSLSLPEYELSSVDYQ 360
QY 478 QOSFYKLOGSVTVTCRNKQWSEPPRCGLDPCVSVSENNKNNIOLKWRNDGKLYAKTGDV 537
Db 361 CQKYYLLKGNKTITCRNGKWSEPTCLHACVIPENIMEAHNIILKWRHTEKIYAHSGEDI 420
QY 538 EFQCKFPFKAMISSPPFRAICQEGKFEYPC 568
Db 421 EFECRGYQKARGLSPFRKTCINGTINYPTC 451

RESULT 10
Q91275
ID Q91275 PRELIMINARY; PRT; 1053 AA.
AC Q91275;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Complement regulatory plasma protein.
OS Parabrax nebulifer (barred sand bass).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Percoidae;
OC Serranidae; Paralabrax.
OX NCBI_TaxID=30873;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE=94318039; PubMed=8042982;
RA Dahmen A., Kaidoh T., Zipfel P.F., Gigli I.;
RT "Cloning and characterization of a cDNA representing a putative
RT complement-regulatory plasma protein from barred sand bass (Parabrax
RT nebulifer).";
RL Biochem. J. 301:391-397(1994).
DR EMBL; L21703; AAA92556.1;
DR HSSP; P08603; IHFH.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 16.
DR SMART; SM00032; CCP; 16.
SQ SEQUENCE 1053 AA; 117597 MW; F27E32C3AD76D503 CRC64;

Query Match 15.0%; Score 476; DB 13; Length 1053;
Best Local Similarity 24.8%; Pred. No. 1.7e-33;
Matches 159; Conservative 94; Mismatches 237; Indels 152; Gaps 34;

QY 37 EDYNPFQVPT--TGEVFIYSEYFNVSFSPKSFTRITCTEEGWSPTPKCL--RMC-SFPPV 92
Db 455 KDPNGFFRGPTGRVLYTYCKDGYKLFTGEWAAEAKCVGVWPELTTCISNTTCGKFPFI 514
QY 93 KNGHSESSGLI--HLEGTQVLIICNTGYSLOKNEKNISCVRGW-----STPPICSTKG 145
Db 515 PNAE-----VIRYPEVQTVVVICNQGYQGAN--SFSCENGWLLYGLSPQICTLRAD 567
QY 146 ECHVPILEANVDAQPK-KESYKVGDLVLFKSRKMLIRV-GSDSVQCYQFGWSPNF----- 198
Db 568 VCGPPPEAENAVVKYSYQREYLSGSEVYLCRDKYPIDEGVDITICRNGQWDEKCTKTS 627
QY 199 -----PTCKGVRCGSP 211
Db 628 CDKLDVDMFTADKEIYEGOTIRYQCLLDAGIATCNNTKWKSPQCK--VPCELP 685
QY 212 POLSNGEVKEIRKEEYGHNEVVEYDCNPNF-IINGPKLIQCVDGWNT-LPTCVQVQVTC 269

```



```

DR PRINTS: PR00453; VNPADOMAIN.
DR ProDom; PD002153; Pentaxin; 1.
DR SMART; SM00032; CCP; 34.
DR SMART; SM00181; EGF; 15.
DR SMART; SM00179; EGF_CA; 9.
DR SMART; SM00001; EGF_like; 3.
DR SMART; SM00159; PTH; 1.
DR SMART; SM00327; VW; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 5.
DR PROSITE; PS00022; EGF_1; UNKNOWN_9.
DR PROSITE; PS01186; EGF_2; 11.
DR PROSITE; PS01187; EGF_CA; 6.
DR PROSITE; PS0234; VWFA; 1.
KW SIGNAL.
FT SIGNAL.
SQ SEQUENCE 3567 AA; 387391 MW; 8FBA8276E12293E5 CRC64;

Query Match 12.4%; Score 393.5; DB 11; Length 3567;
Best Local Similarity 24.8%; Pred. No. 1.6e-25;
Matches 152; Conservative 80; Mismatches 248; Indels 133; Gaps 39;

QY 26 PKIHGFLYDEEDYNPFSQVPTGEVYSC--YNEVSPSKSFWRITCTEEG-WSP-TP 81
DB 1693 PALENGF-YSAEDFH-----AGSTVYQCTSGYLLGDSRMF-----CTDNGSWNGISP 1740
QY 82 KCLRM-----CSPFFV-----KNGHSE 98
DB 1741 SCLDVBECAGSDCSEHASCLNTNGSVVCSNPPYTGDKNCAEPVKRAPENPENGH-- 1798
QY 99 SSGLIHLEGDTVOICNTGYSLONNEKNTSCVERG-WS-TTPCISFTKGECHVPILAN- 155
DB 1799 SSGEITVGVAVTSCDEGHELV-GVSTITCLETGEMDLRPSCEAI--SCGVPPYPENG 1855
QY 156 -VDAQPKESYKVGDKLFSCKRNLRVGSQYQFG-WSPNFTCTCKGOVRSCGPPQ 213
DB 1856 GVDG----SAFTYGVKVVYRCDKGYTLSGDEESACLASGWSHSPVC--ELVKCSQPED 1909
QY 214 LSNGEVKEIRKEY-----GHNEVVEYDCNPFIINGPKKIQC-VDGEWTL-PTVCEQV 266
DB 1910 INNG-----KYILSGLTYSIASYSCENGYSLOQPSLLECTASGSDRAPPSG--QL 1959
QY 267 KTCGYIPELEYGVQVSPYQHGVSVEYCNREYAMIGNMITC-INGIW-TELPWCVA 324
DB 1960 VSCGEPPIKDAVITGS--NFTGNTVAYTCRGYTLAGPDTIVCOANGKSNSSNOCL 2017
QY 325 THOLKCKIAGVNIKTLKLSGKEFNHNSRIRYRCSDFIRYHS---VC-INGKWNPEVD 380
DB 2018 VSCDEPPNVDHASPETAHLFG-----DTAFYICADGYSILADNSOLICNAQGNWYPAG 2071
QY 381 CTEKR--EQFCPPPPPOIP-NAQNMFTTVYQDGEKVAVLCKENYLLPEAKEIVC-KDGRW 436
DB 2072 QAVPRCIAHFCEKPPSVSYSILESVSKAFAAGSVVSKMEGVFLNTSAKIECLRGGEW 2131
QY 437 QSLPRCVES-TAVCGPPPIINGDITSEPLSV-YPPGSTVYVRCQSFYKLOGSVVTCR- 493
DB 2132 SPSPLSQVICPVKCGPPSIANG-----YPSGTNYSEGAVVAVSCHGFKYIKGKSTCEA 2187
QY 494 NKOWSEPPRLCDPCVVYSEENMKNLQKWRNDGKLYAKTGOAVE----FQCKFPKAMI 549
DB 2188 TGOWSKPTPTCHPVCNEPPKVEN-----GFLHTTGTTFESEARFQCPNGYKA-A 2237
QY 550 SSPPPFAICQEGK 562
DB 2238 GSPVF--VCOANR 2248

RESULT 13
Q61405 PRELIMINARY; PRT; 303 AA.
ID Q61405
AC Q61405;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

```

```

DE DE Complement factor H-related protein.
GN A1194696.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90153969; PubMed=1689298;
RA Vlk D.P., Munoz-Canoves P., Kozono H., Martin L.G., Tack B.F.,
RA Chaplin D.D.;
RT "Identification and sequence analysis of 4 complement factor H-related
RT transcripts in mouse liver.";
RL J. Biol. Chem. 265:3193-3201(1990).
DR EMBL; M29007; AAA37413.1; -.
DR HSSP; P10998; 1VVD.
DR MGD; MGI:2138159; A1194696.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 4.
DR SMART; SM00032; CCP; 4.
SQ SEQUENCE 303 AA; 34498 MW; 51C66E50906F4C24 CRC64;

Query Match 12.2%; Score 387.5; DB 11; Length 303;
Best Local Similarity 25.9%; Pred. No. 2.7e-26;
Matches 83; Conservative 38; Mismatches 83; Indels 117; Gaps 4;

QY 1 MLLLFVILISWSTVGGEG----- 20
DB 6 MLLLSNLLTAWLSTAKEGKTCSPPYLNGIYTPHRIIHKSDDEIRYECNYGFYPTGS 65
QY 21 -----TL--CDPPKIHHGLYDEEDYNPFSQVPTGEVYSCYFNVP 62
DB 66 TVSKCTPTGWIPVPRCTLAKPCEFPQKGRLYVEESLRPNFPVSIKNKYSYKCDNGFSPP 125
QY 63 SKSFWRITCTEGERGWSPTKCLRMCSFPFKNGHSESSGLIHLEGTVOICNTGYSLO 122
DB 126 SGYSWDYLKCTAQGWEPVPCVRKCVHYVEVDSTYWEKIYVQGSGLKVCQYNGYSLO 185
QY 123 NEKNISCVRGWSTPTPICSTFKGECHVPILANVDAQPKESYKVGDKLFSCKRNLRV 182
DB 186 GQDTMTCTENGWSPPKPC-----IRI 206
QY 183 GDSVQCYQFGWSPNPTCKGQVRSCGPPQPSNGEVEKEIRKEEYGHNEVYDCNPFI 242
DB 207 NS-----TRTCGPPPIIDNGDITSLSLPEYFLSSVDYQCCKYYL 246
QY 243 INGPKKIQCVGDEWTLTPCV 263
DB 247 LKGNKTIITCRNGWSEPPMCL 267

RESULT 14
Q29528 PRELIMINARY; PRT; 1911 AA.
ID Q29528
AC Q29528;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Complement component receptor type 1 (Fragment).
GN CRI.
OS Papio hamadryas (Hamadryas baboon).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
OC Cercopitheinae; Papio.
OX NCBI_TaxID=9557;
RN [1]
RP SEQUENCE FROM N.A.
RA Clemenza L., Subramanian B.V., Nickells M.W., Hourcade D.E.,
RA Atkinson J.P.;
RT "Primary sequence of the baboon 200 kDa C3b/C4b receptor (CRI).";
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; L39791; AAA62170.1; -.
DR HSSP; P08603; 1HFI.

```

[illegible][illegible]

RESULT 15	
Q29530	
ID	Q29530
PRELIMINARY;	PRT; 2014 AA.
AC	Q29530;
DT	01-NOV-1996 (TREMBLrel. 01; Created)
DT	01-NOV-1996 (TREMBLrel. 01; Last sequence update)
DT	01-DEC-2001 (TREMBLrel. 19; Last annotation update)
DE	Complement receptor 1 (Fragment).
GN	CR1.

Search completed: July 17, 2003, 13:53:00
Job time : 91 secs

GenCore version 5.1.6

Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2003, 13:49:33 ; Search time 24 Seconds
(without alignments)

983.335 Million cell updates/sec

Title: US-09-805-337A-2

Perfect score: 3183

Sequence: 1 MLLFSVLISWSTVGEG.....SSPPFRAICQGRFEYPICE 569

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1538	48.3	1231	1 CFAH_HUMAN	P08603 homo sapien
2	1427.5	44.8	1234	1 CFAH_MOUSE	P06909 mus musculus
3	1050.5	33.0	330	1 FHR1_HUMAN	Q03591 homo sapien
4	927.5	29.1	270	1 FHR2_HUMAN	P36980 homo sapien
5	861	27.0	685	1 CFAH_BOVIN	Q28085 bos taurus
6	783	24.6	668	1 F13B_MOUSE	Q07968 mus musculus
7	764.5	24.0	661	1 F13B_HUMAN	P05160 homo sapien
8	723.5	22.7	330	1 FHR3_HUMAN	Q02985 homo sapien
9	681	21.4	331	1 FHR4_HUMAN	Q92496 homo sapien
10	411	12.9	1033	1 CR2_HUMAN	P20023 homo sapien
11	385	12.1	2039	1 CR1_HUMAN	P17927 homo sapien
12	382	12.0	1025	1 CR2_MOUSE	P19070 mus musculus
13	339.5	10.7	558	1 C4BP_RAT	Q63514 rattus norv
14	328	10.3	610	1 C4BP_BOVIN	Q28065 bos taurus
15	303	9.5	597	1 C4BP_HUMAN	P04003 homo sapien
16	300.5	9.4	469	1 C4BP_MOUSE	P08607 mus musculus
17	281	8.8	830	1 LEM3_HUMAN	P16109 homo sapien
18	274.5	8.6	612	1 LEM2_MOUSE	Q00690 mus musculus
19	264.5	8.3	610	1 LEM2_HUMAN	P16581 homo sapien
20	255	8.0	768	1 LEM3_MOUSE	Q01102 mus musculus
21	253	7.9	768	1 LEM3_RAT	P98106 rattus norv
22	252.5	7.9	769	1 LEM3_SHEEP	P98109 ovis aries
23	247.5	7.8	646	1 LEM3_BOVIN	P42201 bos taurus
24	243.5	7.7	551	1 LEM2_RABIT	P27113 oryctolagus
25	243	7.6	345	1 APOH_MOUSE	Q01339 mus musculus
26	237.5	7.5	345	1 APOH_CANFA	P33703 canis fami
27	235.5	7.4	611	1 LEM2_CANFA	P33730 canis fami
28	233	7.3	549	1 LEM2_RAT	P98105 rattus norv
29	231.5	7.3	263	1 VCP_VACCV	P10998 vaccinia vi
30	231	7.3	345	1 APOH_HUMAN	P02749 homo sapien
31	225.5	7.1	360	1 CCPH_HSVSA	Q01016 herpesvirus
32	224.5	7.1	345	1 APOH_BOVIN	P17690 bos taurus
33	224	7.0	507	1 DAF_CAVPO	Q60401 cavia porce

RESULT 1

ID	CFAH_HUMAN	STANDARD;	PRT;	1231 AA.
AC	P08603: Q14570; P78435; Q9NU86;			
DT	01-AUG-1988 (Rel. 08, Created)			
DT	01-JAN-1990 (Rel. 13, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Complement factor H precursor (H factor 1).			
GN	HF1 OR HF OR CFH.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND VARIANT Y-402.			
RC	TISSUE=Liver;			
RX	MEDLINE=88134059; PubMed=2963625;			
RA	Ripoche J., Day A.J., Harris T.J.R., Sim R.B.;			
RT	"The complete amino acid sequence of human complement factor H.;"			
RL	Biochem. J. 249:593-602(1988).			
RN	[2]			
RP	SEQUENCE OF 53-445 FROM N.A.			
RX	MEDLINE=87054207; PubMed=2946589;			
RA	Schulz T.F., Schwaeble W., Stanley K.K., Weiss E., Dierich M.P.;			
RT	"Human complement factor H: Isolation of cDNA clones and partial cDNA sequence of the 38-kDa tryptic fragment containing the binding site for C3b.;"			
RL	Eur. J. Immunol. 16:1351-1355(1986).			
RN	[3]			
RP	SEQUENCE OF 226-449 FROM N.A., AND PARTIAL SEQUENCE.			
RX	MEDLINE=86169701; PubMed=2937845;			
RA	Kristensen T., Wetzel R.A., Tack B.F.;			
RT	"Structural analysis of human complement protein H: homology with C4b binding protein, beta 2-glycoprotein I, and the Ba fragment of B2.;"			
RL	J. Immunol. 136:3407-3411(1986).			
RN	[4]			
RP	SEQUENCE FROM N.A. (ISOFORM 2).			
RA	Bird C.;			
RT	Submitted (JAN-2000) to the EMBL/GenBank/DBDJ databases.			
RN	[5]			
RP	SEQUENCE OF 1047-1231 FROM N.A.			
RX	MEDLINE=91201892; PubMed=1826708;			
RA	Estaller C., Koistinen V., Schwaeble W., Dierich M.P., Weiss E.H.;			
RT	"Cloning of the 1.4-kb mRNA species of human complement factor H reveals a novel member of the short consensus repeat family related to the carboxy terminal of the classical 150-kDa molecule.;"			
RL	J. Immunol. 146:3190-3196(1991).			
RN	[6]			
RP	SEQUENCE OF 19-35.			
RX	MEDLINE=83048213; PubMed=6215918;			
RA	Sim R.B., Discipio R.G.;			
RT	"Purification and structural studies on the complement-system control protein beta 1H (Factor H).;"			
RL	Biochem. J. 205:285-293(1982).			
RN	[7]			
RP	SEQUENCE OF 1-19 FROM N.A.			

P28175 tachypleus
Q26422 carcinoscor
Q9byh1 homo sapien
P26644 rattus norv
Q09101 drosophila
Q13219 homo sapien
P21115 vaccinia vi
P08174 homo sapien
Q61475 mus musculus
Q01227 vaccinia vi
P24083 vaccinia vi
P49457 pongo pygma

FT	DOMAIN	689	744	SUSHI 12.
FT	DOMAIN	751	803	SUSHI 13.
FT	DOMAIN	807	862	SUSHI 14.
FT	DOMAIN	866	930	SUSHI 15.
FT	DOMAIN	935	992	SUSHI 16.
FT	DOMAIN	993	1049	SUSHI 17.
FT	DOMAIN	1052	1108	SUSHI 18.
FT	DOMAIN	1113	1169	SUSHI 19.
FT	DOMAIN	1171	1234	SUSHI 20.
FT	DISULFID	21	66	BY SIMILARITY.
FT	DISULFID	52	80	BY SIMILARITY.
FT	DISULFID	85	129	BY SIMILARITY.
FT	DISULFID	114	141	BY SIMILARITY.
FT	DISULFID	146	192	BY SIMILARITY.
FT	DISULFID	178	205	BY SIMILARITY.
FT	DISULFID	210	251	BY SIMILARITY.
FT	DISULFID	237	262	BY SIMILARITY.
FT	DISULFID	267	309	BY SIMILARITY.
FT	DISULFID	294	320	BY SIMILARITY.
FT	DISULFID	325	374	BY SIMILARITY.
FT	DISULFID	357	385	BY SIMILARITY.
FT	DISULFID	389	431	BY SIMILARITY.
FT	DISULFID	416	442	BY SIMILARITY.
FT	DISULFID	448	494	BY SIMILARITY.
FT	DISULFID	477	505	BY SIMILARITY.
FT	DISULFID	509	553	BY SIMILARITY.
FT	DISULFID	536	564	BY SIMILARITY.
FT	DISULFID	569	610	BY SIMILARITY.
FT	DISULFID	597	622	BY SIMILARITY.
FT	DISULFID	629	672	BY SIMILARITY.
FT	DISULFID	658	683	BY SIMILARITY.
FT	DISULFID	690	732	BY SIMILARITY.
FT	DISULFID	718	743	BY SIMILARITY.
FT	DISULFID	752	791	BY SIMILARITY.
FT	DISULFID	780	802	BY SIMILARITY.
FT	DISULFID	808	850	BY SIMILARITY.
FT	DISULFID	836	861	BY SIMILARITY.
FT	DISULFID	867	920	BY SIMILARITY.
FT	DISULFID	906	931	BY SIMILARITY.
FT	DISULFID	936	978	BY SIMILARITY.
FT	DISULFID	964	989	BY SIMILARITY.
FT	DISULFID	994	1037	BY SIMILARITY.
FT	DISULFID	1023	1048	BY SIMILARITY.
FT	DISULFID	1053	1096	BY SIMILARITY.
FT	DISULFID	1082	1107	BY SIMILARITY.
FT	DISULFID	1114	1157	BY SIMILARITY.
FT	DISULFID	1143	1168	BY SIMILARITY.
FT	DISULFID	1172	1223	BY SIMILARITY.
FT	DISULFID	1206	1233	BY SIMILARITY.
FT	CARBOHYD	676	676	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	721	721	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	773	773	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	801	801	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1030	1030	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1061	1061	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1225	1225	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1234	1234	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	SEQUENCE	1234	1234	AA; 139082 MW; C5AC02F341B957F7 CRC64;
FT	Query Match		44.8%;	Score 1427.5; DB 1; Length 1234;
FT	Best Local Similarity		31.0%;	Pred. No. 3.4e-97;
FT	Matches	283;	Conservative	82; Mismatches 178; Indels 369; Gaps 5;
QY	23	CDFPKTHGLFLDEEDYNFSPQVTEVFYSCYENFVSPKSFWRITCTEENGSPTPK 82		
Db	325	CEFPQPKYGLYEESLRNFPVSGNKYKCDNGFSPPGYSWDYLCTAAGWEPEVP 384		
QY	83	CLRMCSFPVKNKSESSGLHLECDTVQICNTGYSLONNEKNISCVRGCHSTPPIC-- 140		
Db	385	CVKCKVFHYVENGDSAYWEKVYQGOSLKVQCTNGYSLONGQDTMTCTENGWSPPKCTR 444		
QY	141	-----SFTKG----- 145		
Db	445	IKTCSADTHIDNGLFLESSSIYALNRETYSRCKQGYVTNTGELSGSITCLONGHSPQS 504		

QY	141	-----SFTKG----- 145
Db	505	CIKSCDMPVPENSITKNTRTWFKLNDKLDYECLVGFENYKHTKSTCTCTYYGHSPTSC 564
QY	146	---ECHVPILBANVDAQPKKESYKGVDLAFSPKRNLIIRVGSVDSVQCYQFGWSPNPTCK 202
Db	565	YERECVPTLDRKLWSPRKEKYRVGDLLEFSCHS--HRVGPDSVQCYHFQWSPGFPPTCK 623
QY	203	GOVRSQPPQLSNGEYKTRKEEYGHNEVVEYDCNPNFLINGPKKIQCVGEGWTTLPCTC 262
Db	624	GOVASCAPPEILNGEINGAKKVEYSHGEVYKVDKPRFLKLGPNKIQCVGDNWTTLPVC 683
QY	263	VEQVKTCGYIPELEYGYVQSPVPPYQHGVSVEVNCREYAMIGNNMITCINGIWTPLMC 322
Db	684	IEBERTCGDIPLEHGSASAKSVPPYHHGDSVEFICEENFTMIGHGSVCSIGKWTQPKC 743
QY	323	VATHQLKCKIAGVNIKTLKSGKFNHNSRIYRCSDFIRYHSHVCINGKNWPEVDC 382
Db	744	VATDQLEKCRVLKSTGIEAIKPKLTFTHTNSTDYKCRDKQEYERSICINGKWDPEPNC 803
QY	383	EKREQFCPPPPQLPNAGMTTNYQDGKAVLCKENYLLPEAKETVCKDGRWQSLPRC 442
Db	804	SKYS--CPPPPQIPNTQVIETTVYLDGEKSLVLCQNYLTQDSEENVCKDGRWQSLPRC 861
QY	443	VE----- 444
Db	862	IEKIPCSQPTIEHGSINLPRSSERRDSIESSSHEHGTFTSYVCDGFPRIPEENRITCY 921
QY	445	----- 444
Db	922	MGKSTPPRCVGLPCGPPPSPIPIGTVSLESEYQHGEEVYHCGSTGFGIDGPAFIIEGG 981
QY	445	----- 444
Db	982	KWSDPPKCTKDCDVLPTVKNAILRCKSKSYRTGEQVFRCSQPYQMNGSDVTVCVNSR 1041
QY	445	----- 444
Db	1042	WIGQPVCKDNCSDVPPHPNATIVTRTKNKLHGDRVRYECNKPFLFGQVYVNCENGW 1101
QY	445	-----STAYCGPPPSINNGDTTSFPLSVYPPGTVTYRCSQPYKLOGSVTVTCRNQ 496
Db	1102	TEKPKCRDSTGKCGPPPDINGDITSLSLPVVEPLSVSEYQCYKYLKGGKKTITCTNGK 1161
QY	497	WSEPPRCCLDPVYSEENMKNNIQKWRNDGKLYAKTGDAVEQCKFPFKAMISSPFFRA 556
Db	1162	WSEPPFCLHACVIPENIMESHNIILKWRHTEKIYSHSGDIEFGCKYGYKARDSPFPRT 1221
QY	557	ICOEGKFEPYIC 568
Db	1222	KCINGTINPTC 1233

RESULT 3		
FHRL_HUMAN	STANDARD;	PRT; 330 AA.
ID	FHRL_HUMAN	Q03591; Q9UJ17;
AC	Q03591; Q9UJ17;	
DT	01-JUN-1994 (Rel. 29, Created)	
DT	01-JUN-1994 (Rel. 29, Last sequence update)	
DT	15-JUN-2002 (Rel. 41, Last annotation update)	
DE	Complement factor H-related protein 1 precursor (FHR-1) (H factor-like protein 1) (H-factor like 1) (H36).	
DE		
GN	HFL1 OR FHRL1 OR CFHL.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OX	NCBI_Taxid=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=Liver;	
RC	MEDLINE=91201892; PubMed=1826708;	
RA	Estaller C., Kolstinen V., Schwaebel W., Dierich M.P., Weiss E.H.;	

RT "Cloning of the 1.4-kb mRNA species of human complement factor H reveals a novel member of the short consensus repeat family related to the carboxy terminal of the classical 150-kDa molecule."; J. Immunol. 146:3190-3196(1991).

RL [2]

RN SEQUENCE FROM N.A.

RA MEDLINE=20245597; PubMed=10781834;

RX Male D.A., Ormsby R.J., Ranganathan S., Giannakis E., Gordon D.L.;

RT "Complement factor H: sequence analysis of 221 kb of human genomic DNA containing the entire FH, fHR-1 and fHR-3 genes.";

RL Mol. Immunol. 37:41-52(2000).

RN [3]

RN SEQUENCE FROM N.A.

RA Hall R.;

RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.

RN [4]

RN SEQUENCE OF 4-330 FROM N.A.

RX MEDLINE=9123700; PubMed=1825108;

RX Skerka C., Horstmann R.D., Zipfel P.F.;

RA "Molecular cloning of a human serum protein structurally related to complement factor H.";

RL J. Biol. Chem. 266:12015-12020(1991).

RN [5]

RN STRUCTURE OF CARBOHYDRATES.

RX MEDLINE=9123700; PubMed=1825108;

RA Timmann C., Leippe M., Horstmann R.D.;

RT "Two major serum components antigenically related to complement factor H are different glycosylation forms of a single protein with no factor H-like complement regulatory functions.";

RL J. Immunol. 146:1265-1270(1991).

RN [6]

RN REVIEW.

RX MEDLINE=94226679; PubMed=8172644;

RA Zipfel P.F., Skerka C.;

RT "Complement factor H and related proteins: an expanding family of complement-regulatory proteins?";

RL Immunol. Today 15:121-126(1994).

CC -!- FUNCTION: MIGHT BE INVOLVED IN COMPLEMENT REGULATION. CAN ASSOCIATE WITH LIPOPROTEINS AND MAY PLAY A ROLE IN LIPID METABOLISM.

CC -!- SUBCELLULAR LOCATION: Extracellular.

CC -!- TISSUE SPECIFICITY: LIVER.

CC -!- PTM: N-GLYCOSYLATED. TWO FORMS ARE OBSERVED; ONE WITH A SINGLE SIDE-CHAIN AND THE OTHER WITH TWO.

CC -!- SIMILARITY: CONTAINS 5 SUSHI (SCR) DOMAINS.

CC -!- SIMILARITY: STRONG. TO FACTOR H.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; M65292; AAA35946.1; -;

DR EMBL; M65293; AAA35947.1; -;

DR EMBL; AL049741; CAB53063.1; -;

DR EMBL; X56209; CAA39666.1; -;

DR PIR; A40455; A40455.

DR PIR; S14604; S14604.

DR HSSP; P10998; 1YVD.

DR Genew; HGNC:4888; HFL1.

DR MIM; 134371; -;

DR InterPro; IPR000436; Sushi_SCR_CCP.

DR Pfam; PF00084; sushi; 5.

DR SMART; SM00032; CCP; 5.

KW Repeat; Glycoprotein; Sushi; Signal; Polymorphism.

FT SIGNAL 1 18

FT CHAIN 19 330 COMPLEMENT FACTOR H-RELATED PROTEIN 1.

FT DOMAIN 22 84 SUSHI 1.

FT DOMAIN 86 141 SUSHI 2.

FT DOMAIN 146 202 SUSHI 3.

FT	DOMAIN	207	263	SUSHI 4.
FT	DOMAIN	265	328	SUSHI 5.
FT	DISULFID	23	72	BY SIMILARITY.
FT	DISULFID	55	83	BY SIMILARITY.
FT	DISULFID	87	129	BY SIMILARITY.
FT	DISULFID	114	140	BY SIMILARITY.
FT	DISULFID	147	190	BY SIMILARITY.
FT	DISULFID	176	201	BY SIMILARITY.
FT	DISULFID	208	251	BY SIMILARITY.
FT	DISULFID	237	262	BY SIMILARITY.
FT	DISULFID	266	317	BY SIMILARITY.
FT	DISULFID	300	327	BY SIMILARITY.
FT	CARBOHYD	126	126	N-LINKED (GLCNAC. .).
FT	CARBOHYD	194	194	N-LINKED (GLCNAC. .).
FT	VARIANT	157	157	Y -> H.
FT	VARIANT	159	159	/FTid=VAR_001980.
FT	VARIANT	175	175	/FTid=VAR_001981.
FT	VARIANT	175	175	O -> E.
FT	CONFLICT	71	71	T -> N (IN REF. 4).
SQ	SEQUENCE	330 AA;	37661 MW;	8DC0D3F92A85E035 CRC64;

Query Match 33.08; Score 1050.5; DB 1; Length 330;
 Best Local Similarity 38.08; Pred. No. 3.6e-70;
 Matches 216; Conservative 30; Mismatches 81; Indels 241; Gaps 5;

QY	1	MLLLSVILISWSVTGEGTLCDFPKIHGHFLYDEEDYNPFPSOVPTGEVFFYSCEYNFV	60
DB <td>1 <td>MLLLSVILISWSVTGEGTLCDFPKIHGHFLYDEEDYNPFPSOVPTGEVFFYSCEYNFV <td>60</td> </td></td>	1 <td>MLLLSVILISWSVTGEGTLCDFPKIHGHFLYDEEDYNPFPSOVPTGEVFFYSCEYNFV <td>60</td> </td>	MLLLSVILISWSVTGEGTLCDFPKIHGHFLYDEEDYNPFPSOVPTGEVFFYSCEYNFV <td>60</td>	60
QY <td>61 <td>SPKSFWRITCTEESGWSPTPKCLRMCSFPVKNHSESSGLIHLEGGTVOIICNTGYSL</td> <td>120</td> </td>	61 <td>SPKSFWRITCTEESGWSPTPKCLRMCSFPVKNHSESSGLIHLEGGTVOIICNTGYSL</td> <td>120</td>	SPKSFWRITCTEESGWSPTPKCLRMCSFPVKNHSESSGLIHLEGGTVOIICNTGYSL	120
DB <td>61 <td>SPKSFWRITCTEESGWSPTPKCLRMCSFPVKNHSESSGLIHLEGGTVOIICNTGYSL</td> <td>120</td> </td>	61 <td>SPKSFWRITCTEESGWSPTPKCLRMCSFPVKNHSESSGLIHLEGGTVOIICNTGYSL</td> <td>120</td>	SPKSFWRITCTEESGWSPTPKCLRMCSFPVKNHSESSGLIHLEGGTVOIICNTGYSL	120
QY <td>121 <td>ONNEKNISCVERGWSPTPKCSFTTKGCHVPILEANVDAQPKESYKVGVDVLKSCRNLI</td> <td>180</td> </td>	121 <td>ONNEKNISCVERGWSPTPKCSFTTKGCHVPILEANVDAQPKESYKVGVDVLKSCRNLI</td> <td>180</td>	ONNEKNISCVERGWSPTPKCSFTTKGCHVPILEANVDAQPKESYKVGVDVLKSCRNLI	180
DB <td>121 <td>ONNEKNISCVERGWSPTPKCSFTTKGCHVPILEANVDAQPKESYKVGVDVLKSCRNLI</td> <td>180</td> </td>	121 <td>ONNEKNISCVERGWSPTPKCSFTTKGCHVPILEANVDAQPKESYKVGVDVLKSCRNLI</td> <td>180</td>	ONNEKNISCVERGWSPTPKCSFTTKGCHVPILEANVDAQPKESYKVGVDVLKSCRNLI	180
QY <td>181 <td>RVGSDSVQCYQFGWSPNFPCTKGVRSCTGPPQSLNNGVKEIRKEEYGHNEVWEYDCNPN</td> <td>240</td> </td>	181 <td>RVGSDSVQCYQFGWSPNFPCTKGVRSCTGPPQSLNNGVKEIRKEEYGHNEVWEYDCNPN</td> <td>240</td>	RVGSDSVQCYQFGWSPNFPCTKGVRSCTGPPQSLNNGVKEIRKEEYGHNEVWEYDCNPN	240
DB <td>181 <td>RVGSDSVQCYQFGWSPNFPCTKGVRSCTGPPQSLNNGVKEIRKEEYGHNEVWEYDCNPN</td> <td>240</td> </td>	181 <td>RVGSDSVQCYQFGWSPNFPCTKGVRSCTGPPQSLNNGVKEIRKEEYGHNEVWEYDCNPN</td> <td>240</td>	RVGSDSVQCYQFGWSPNFPCTKGVRSCTGPPQSLNNGVKEIRKEEYGHNEVWEYDCNPN	240
QY <td>241 <td>FIINGPKKIOCVGDGWTTLPTCVBQVKTCGYIPELEYGYVQSPVPYQHGVSVEVNCNRNE</td> <td>300</td> </td>	241 <td>FIINGPKKIOCVGDGWTTLPTCVBQVKTCGYIPELEYGYVQSPVPYQHGVSVEVNCNRNE</td> <td>300</td>	FIINGPKKIOCVGDGWTTLPTCVBQVKTCGYIPELEYGYVQSPVPYQHGVSVEVNCNRNE	300
DB <td>241 <td>FIINGPKKIOCVGDGWTTLPTCVBQVKTCGYIPELEYGYVQSPVPYQHGVSVEVNCNRNE</td> <td>300</td> </td>	241 <td>FIINGPKKIOCVGDGWTTLPTCVBQVKTCGYIPELEYGYVQSPVPYQHGVSVEVNCNRNE</td> <td>300</td>	FIINGPKKIOCVGDGWTTLPTCVBQVKTCGYIPELEYGYVQSPVPYQHGVSVEVNCNRNE	300
QY <td>361 <td>DIERYSHVSCINGKNWPEVDCTEKREQFCPPPPQIPNAPNAQNTTTVNYQDGEKVAVLCKEN</td> <td>420</td> </td>	361 <td>DIERYSHVSCINGKNWPEVDCTEKREQFCPPPPQIPNAPNAQNTTTVNYQDGEKVAVLCKEN</td> <td>420</td>	DIERYSHVSCINGKNWPEVDCTEKREQFCPPPPQIPNAPNAQNTTTVNYQDGEKVAVLCKEN	420
DB <td>361 <td>DIERYSHVSCINGKNWPEVDCTEKREQFCPPPPQIPNAPNAQNTTTVNYQDGEKVAVLCKEN</td> <td>420</td> </td>	361 <td>DIERYSHVSCINGKNWPEVDCTEKREQFCPPPPQIPNAPNAQNTTTVNYQDGEKVAVLCKEN</td> <td>420</td>	DIERYSHVSCINGKNWPEVDCTEKREQFCPPPPQIPNAPNAQNTTTVNYQDGEKVAVLCKEN	420
QY <td>421 <td>YLLPEAKEIVCKDGRWQSLPRCVSTAYCGPPPSINNGDTSFPLSVYPPGSTVYRQCS</td> <td>480</td> </td>	421 <td>YLLPEAKEIVCKDGRWQSLPRCVSTAYCGPPPSINNGDTSFPLSVYPPGSTVYRQCS</td> <td>480</td>	YLLPEAKEIVCKDGRWQSLPRCVSTAYCGPPPSINNGDTSFPLSVYPPGSTVYRQCS	480
DB <td>421 <td>YLLPEAKEIVCKDGRWQSLPRCVSTAYCGPPPSINNGDTSFPLSVYPPGSTVYRQCS</td> <td>480</td> </td>	421 <td>YLLPEAKEIVCKDGRWQSLPRCVSTAYCGPPPSINNGDTSFPLSVYPPGSTVYRQCS</td> <td>480</td>	YLLPEAKEIVCKDGRWQSLPRCVSTAYCGPPPSINNGDTSFPLSVYPPGSTVYRQCS	480
QY <td>541 <td>CKFPHKAMISSPPFRATICEGKEEYVPC</td> <td>568</td> </td>	541 <td>CKFPHKAMISSPPFRATICEGKEEYVPC</td> <td>568</td>	CKFPHKAMISSPPFRATICEGKEEYVPC	568
DB <td>541 <td>CKFPHKAMISSPPFRATICEGKEEYVPC</td> <td>568</td> </td>	541 <td>CKFPHKAMISSPPFRATICEGKEEYVPC</td> <td>568</td>	CKFPHKAMISSPPFRATICEGKEEYVPC	568
QY <td>300 <td>CKRGYLRSSRSHLRTTCWDGKLEYPTC</td> <td>327</td> </td>	300 <td>CKRGYLRSSRSHLRTTCWDGKLEYPTC</td> <td>327</td>	CKRGYLRSSRSHLRTTCWDGKLEYPTC	327
DB <td>300 <td>CKRGYLRSSRSHLRTTCWDGKLEYPTC</td> <td>327</td> </td>	300 <td>CKRGYLRSSRSHLRTTCWDGKLEYPTC</td> <td>327</td>	CKRGYLRSSRSHLRTTCWDGKLEYPTC	327

RESULT 4
 FHR2_HUMAN
 ID FHR2_HUMAN STANDARD; PRT; 270 AA.
 AC P36980; Q14310;
 DT 01-JUN-1994 (Rel. 29, Created)

01-JUN-1994 (Rel. 29, Last sequence update)
 15-JUN-2002 (Rel. 41, Last annotation update)
 Complement factor H-related protein 2 precursor (FHR-2) (H factor-like protein 2) (H factor-like 3) (DDESK59).
 HFL3 OR FHR2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_TaxID=9606;
 [1]
 SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=92251200; PubMed=1533657;
 RA Skerka C., Ilmanen C., Horstmann R.D., Zipfel P.E.;
 RT "Two additional human serum proteins structurally related to complement factor H. Evidence for a family of factor H-related genes.";
 RL J. Immunol. 148:3313-3318(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95402981; PubMed=7672821;
 RA Skerka C., Moulds J.M., Taillon-Miller P., Hourcade D., Zipfel P.F.;
 RT "The human factor H-related gene 2 (FHR2): structure and linkage to the coagulation factor XIIIb gene.";
 RL Immunogenetics 42:268-274(1995).
 RN [3]
 RP REVIEW.
 RX MEDLINE=94226679; PubMed=8172644;
 RA Zipfel P.F., Skerka C.;
 RT "Complement factor H and related proteins: an expanding family of complement-regulatory proteins?";
 RL Immunol. Today 15:121-126(1994).
 CC -1- FUNCTION: MIGHT BE INVOLVED IN COMPLEMENT REGULATION. CAN ASSOCIATE WITH LIPOPROTEINS AND MAY PLAY A ROLE IN LIPID METABOLISM.
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A TRUNCATED FORM; MAY BE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- PTM: N-GLYCOSYLATED. TWO FORMS ARE OBSERVED; ONE WITH A SINGLE SIDE CHAIN, THE OTHER IS NOT GLYCOSYLATED.
 CC -1- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
 CC -1- SIMILARITY: STRONG, TO FACTOR H.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 DR EMBL; X64877; CAA46096.1; -;
 DR EMBL; X86564; CAA60375.1; -;
 DR EMBL; X86565; CAA60375.1; JOINED.
 DR EMBL; X86566; CAA60375.1; JOINED.
 DR EMBL; X86567; CAA60375.1; JOINED.
 DR PIR; S24564; S24564.
 DR KSSP; P08603; IHFI.
 DR Genew; HGNC:4890; HFL3.
 DR MIM; 600889; -;
 DR InterPro; IPR000436; Sush1_SCR_CCP.
 DR Pfam; PF00084; sush1; 4.
 DR SMART; SM00032; CCP; 4.
 KW Repeat; Glycoprotein; Sush1; Signal; Alternative splicing.
 FT SIGNAL 1 18
 FT CHAIN 19 270
 FT DOMAIN 22 84
 FT DOMAIN 86 141
 FT DOMAIN 148 204
 FT DOMAIN 206 268
 FT DISULFID 23 72
 FT DISULFID 55 83
 FT DISULFID 87 129
 BY SIMILARITY.

FT DISULFID 114 140 BY SIMILARITY.
 FT DISULFID 149 192 BY SIMILARITY.
 FT DISULFID 178 203 BY SIMILARITY.
 FT DISULFID 207 257 BY SIMILARITY.
 FT DISULFID 241 267 BY SIMILARITY.
 FT CARBOHYD 126 126 N-LINKED (GLCNAC...).
 FT VARSPIC 144 171 ISAEKCGPPPIDNGDITSPLLSYAPG -> S (IN TRUNCATED ISOFORM).
 FT CONFLICT 85 85 R -> K (IN REF. 2).
 SQ SEQUENCE 270 AA; 30651 MW; ELB2E9F139B217A8 CRC64;
 Query Match 29.1%; Score 927.5; DB 1; Length 270;
 Best Local Similarity 34.6%; Pred. No. 3.1e-61;
 Matches 198; Conservative 20; Mismatches 47; Indels 307; Gaps 3;
 QY 1 MLLFSLVILISWSTVGGEGTLCDFPKIHGHFLYDEEDYNNPFSQVPTGEVFFYSCEYNFV 60
 DB 1 MLLFSLVILISWSTVGGEGTLCDFPKIHGHFLYDEEDYNNPFSQVPTGEVFFYSCEYNFV 60
 QY 61 SPSKSFWRITCTTEEGWSPPTKCLRMCSFPFVKNGHSESLIHLGDTVQIICNTGYSL 120
 DB 61 SPSKSFWRITCTTEEGWSPPTKCLRMCSFPFVKNGHSESLIHLGDTVQIICNTGYSL 120
 QY 121 QNNEKNISCVERGWSPTTCSFTKGECHVPLEANVDAQPKKESYKVGVDVLFKFSRKNLI 180
 DB 121 QNNEKNISCVERGWSPTTCSFTKGECHVPLEANVDAQPKKESYKVGVDVLFKFSRKNLI 180
 QY 181 RVGSDSVQCVQFGWSPNFTCKQVRSRGGPPQLSNCEYKEIRKEEYGHNEVVEYDCNPN 240
 DB 141 - - - - - 140
 QY 241 FIINGPKIQVDGWTTLPTCQEYQVKTGCIPELEYGVQPSVPPYQHGVSVEVACRNE 300
 DB 141 - - - - - 140
 QY 301 YAMIGNMTICINGIWTCLPMCVATHQLRCKTAGVNIKTLKLSKEFNHNSRIYRCS 360
 DB 141 - - - - - 140
 QY 361 DIPRYRHSVINGKNWPEVDCTEKREQFCPPPPPIPAQNMNTTVNYQGEKVAVLCKEN 420
 DB 141 - - - - - 140
 QY 421 YLLPEAKEIVCKDRWQSLPRCVESYACGPPPSINNGDTSPLSVYPPGSTVYRCOS 480
 DB 141 - - - - - RSTISAEKCGPPPIDNGDITSPLLSYAPGSSVEYQCCN 180
 QY 481 FYKLGQSVTVTCRNKQWSEPPRCCLDPVCVYSEENNNKNTOLKWRNDGKLYAKTGDAVEFQ 540
 DB 181 LYOLEGNNQITCRNGQWSEPPRCCLDPVCVYSEENNNKNTOLKWRNDGKLYAKTGDAVEFQ 540
 QY 541 CK---FPHKAMISSPPFRAICQEGKFEYPICE 569
 DB 241 CKSGVHPTK-----SHSFRAMCQNGKLVYPSCE 268
 RESULT 5
 CFAH_BOVIN STANDARD; PRT; 685 AA.
 ID CFAH_BOVIN
 AC Q28085;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Complement factor H (H factor 1) (Fragments).
 GN HFI.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.
 NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE OF 17-685 FROM N.A., AND SEQUENCE OF 1-16.
 RC TISSUE=Liver;

KW Plasma: Blood coagulation; Repeat; Glycoprotein; Signal; Sushi.
 FT SIGNAL 1 20 BY SIMILARITY.
 FT CHAIN 21 668 COAGULATION FACTOR XIII B CHAIN.
 FT DOMAIN 24 88 SUSHI 1.
 FT DOMAIN 90 147 SUSHI 2.
 FT DOMAIN 152 209 SUSHI 3.
 FT DOMAIN 212 268 SUSHI 4.
 FT DOMAIN 273 328 SUSHI 5.
 FT DOMAIN 335 390 SUSHI 6.
 FT DOMAIN 395 451 SUSHI 7.
 FT DOMAIN 453 516 SUSHI 8.
 FT DOMAIN 523 579 SUSHI 9.
 FT DOMAIN 581 647 SUSHI 10.
 FT DISULFID 25 76 BY SIMILARITY.
 FT DISULFID 59 87 BY SIMILARITY.
 FT DISULFID 91 135 BY SIMILARITY.
 FT DISULFID 118 146 BY SIMILARITY.
 FT DISULFID 153 197 BY SIMILARITY.
 FT DISULFID 180 208 BY SIMILARITY.
 FT DISULFID 213 255 BY SIMILARITY.
 FT DISULFID 241 267 BY SIMILARITY.
 FT DISULFID 274 316 BY SIMILARITY.
 FT DISULFID 302 327 BY SIMILARITY.
 FT DISULFID 336 378 BY SIMILARITY.
 FT DISULFID 364 389 BY SIMILARITY.
 FT DISULFID 396 439 BY SIMILARITY.
 FT DISULFID 425 450 BY SIMILARITY.
 FT DISULFID 454 505 BY SIMILARITY.
 FT DISULFID 486 515 BY SIMILARITY.
 FT DISULFID 524 567 BY SIMILARITY.
 FT DISULFID 553 578 BY SIMILARITY.
 FT DISULFID 582 636 BY SIMILARITY.
 FT DISULFID 616 646 BY SIMILARITY.
 FT CARBOHYD 162 162 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 545 545 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 668 AA; 80BC9E00A953FA6 CRC64;

Query Match 24.6%; Score 783; DB 1; Length 668;
 Best Local Similarity 28.7%; Pred. No. 3.6e-50;
 Matches 187; Conservative 95; Mismatches 266; Indels 104; Gaps 18;

QY 7 VILISWVTVGGRTCLDFPKIHG---FLYD-EEDYFNPQVPTGEVFFYSCYFNVS 61
 DB 9 ILLLLSGELAYAEKQCDPTVNGRIAYQYTFKSFYFPMs---VDKLSFFCLAGYAT 65
 QY 62 PSKSFWRITCTEGNSPTPKLURMCSFPVKNHSESSLIHLEGDTVQIICNTGYSLQ 121
 DB 66 ESGKQEEQIRCTAEGNSPNRCVKCLKPLDRNGYVNDKVLKQLGRMSYGCSSGYKT 125
 QY 122 --NNEKNISCVRGWSTPFCSTKGECHVPLE-----ANVDAQPKESYKGVDLKFSCKNL 179
 DB 126 GGDKEEVVHCLSGNSQSPSCRKEQETCLAPELHGNYSPTORTFKVDIVAYTCTAGYY 185
 QY 154 -----ANVDAQPKESYKGVDLKFSCKNL 179
 DB 186 TTTGKOTGAECQANCWSLTPQCNKLMCSRLIENGYFHPVQTYEEDGLVQFFCHENY 245
 QY 180 IRVGSVVOCYQFGSNPFTCKGVQRSGPPQLSNGEVKKEIKKEEYGHNEVEVDGCP 239
 DB 246 YLGSGLIQCYNFGWYSPESICBGRNRNCPVPVPLNSKIOP-HSTYRHRGERVHCECL 304
 QY 240 NFINGPKIQVDGWTTLPTCQEVOK--TCGYIPELEYGVQVSPVPYHGVSVYVNC 297
 DB 305 NFVIOGSELLCNGKWTPEPCIEBEKVACEQPPSVENGVAHPHSEIYYSQDKVYRC 364
 QY 298 RNEYAMIGNMTICINGIWTCLPMCVATHOLARCK---IA-GVNIKTLKLSGKGFNN 352
 DB 365 GGGYSLRGSGSTTCNRRGRTLPPECV--ENIENCKPPDPIANGVVVDGLL-----ASYTGG 418
 QY 353 SRIKRCSDIFRYHSV---CINGKNWPEVDCTEKREQCPPPPTGPNAAQNMTTTVNQ- 408
 DB 419 SSVEVRCNEYLLKGSCTSRCEGAWSSPPVCLP-----CTIDVDMNRNRLQKWKYEG 474

QY 409 ---DGEKVAVLCKENYLLPEA---KEIVCKDGRWQ-SLPRCV--ESTAYGPPPSINNGD 459
 DB 475 KILHGLDIDFVQKQYGLNSPLSEISAQCNRGDVRYPWCKIRKSKGMCASPPVIRNGD 534
 QY 460 TTFPPLSVYPPGTVTYRCOSFYKLGQSVTVTCRNKOWSEPPRCRLDPCVYVEENNNKNI 519
 DB 535 IVSSAARTYENGSSVEYRCFNDHFLQGSQNVYCDGVMTTPPSCLEPCTLSFVEMDKNYL 594
 QY 520 QLKWRNDGKLYAKTGDVAVEFOCK---PPHKAMISPPFPRAICQEGKFEYFIC 568
 DB 595 QLKWNFDNRPLILHGEVIEFMCCKRDAYISETSAGSVLRVQCDRGLKYPKC 646

RESULT 7

F13B_HUMAN
 ID F13B_HUMAN STANDARD; PRT; 661 AA.
 AC P05160;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Coagulation factor XIII B chain precursor (Protein-glutamine gamma-glutamyltransferase B chain) (Transglutaminase B chain) (Fibrin stabilizing factor B subunit).
 GN F13B.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-91105034; PubMed-2271707;
 RA Bottene R.E., Ichinose A., Davie E.W.;
 RT "Nucleotide sequence of the gene for the b subunit of human factor XIII.";
 RL Biochemistry 29:11195-11209(1990).
 RN [2]
 RP SEQUENCE OF 2-661 FROM N.A.
 RX MEDLINE-87026535; PubMed-3021194;
 RA Ichinose A., McMullen B.A., Fujikawa K., Davie E.W.;
 RT "Amino acid sequence of the b subunit of human factor XIII, a protein composed of ten repetitive segments.";
 RL Biochemistry 25:4633-4638(1986).
 RN [3]
 RP REVISIONS.
 RA Ichinose A.;
 RL Submitted (FEB-1987) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 1-20 FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE-90251467; PubMed-2339067;
 RA Grundmann U., Nerlich C., Rein T., Zettlmeissl G.;
 RT "Complete cDNA sequence encoding the B subunit of human factor XIII.";
 RL Nucleic Acids Res. 18:2817-2817(1990).
 RN [5]
 RP VARIANT PHE-450.
 RX MEDLINE-8324218; PubMed-8324218;
 RA Hashiguchi T., Saito M., Morishita E., Matsuda T., Ichinose A.;
 RT "Two genetic defects in a patient with complete deficiency of the b-subunit for coagulation factor XIII.";
 RL Blood 82:145-150(1993).
 CC -!- FUNCTION: THE B CHAIN OF FACTOR XIII IS NOT CATALYTICALLY ACTIVE, BUT IS THOUGHT TO STABILIZE THE A SUBUNITS AND REGULATE THE RATE OF TRANSLUTAMINASE FORMATION BY THROMBIN.
 CC -!- SUBUNIT: TETRAMER OF TWO A CHAINS AND TWO B CHAINS.
 CC -!- DISEASE: Defects in F13B can result in a lifelong bleeding tendency, defective wound healing, and habitual abortion.
 CC -!- SIMILARITY: CONTAINS 10 SUSHI (SCR) DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

entities requires a license agreement (See http://www.isb-sib.ch/announcement
or send an email to license@isb-sib.ch).
-----
CC      EMBL; M64554; AAA51821.1; ALT_SEQ.
CC      EMBL; M14057; AAA88042.1; -.
CC      EMBL; X51823; CAA36123.1; -.
DR      PIR; A23830; A23830.
DR      PIR; A36397; A36397.
DR      PIR; S09980; S09980.
DR      HSSP; P08603; 1HFI.
DR      Genew; HGNC:3534; F13B.
DR      MTM; 134580; -.
DR      InterPro; IPR000436; Sushi_SCR_CCP.
DR      Pfam; PF00084; sushi; 8.
DR      SMART; SM00032; CCP; 8.
KW      Plasma; Blood coagulation; Repeat; Glycoprotein; Signal; Sushi;
KW      Disease mutation.
FT      SIGNAL          1      20
FT      CHAIN           21      661
FT      DOMAIN          24      88
FT      DOMAIN          90      147
FT      DOMAIN          90      147
FT      DOMAIN          152      209
FT      DOMAIN          212      268
FT      DOMAIN          273      328
FT      DOMAIN          335      390
FT      DOMAIN          395      451
FT      DOMAIN          453      516
FT      DOMAIN          523      579
FT      DOMAIN          581      647
FT      DISULFID        25      76
FT      DISULFID        59      87
FT      DISULFID        91      135
FT      DISULFID        118      146
FT      DISULFID        153      197
FT      DISULFID        180      208
FT      DISULFID        213      255
FT      DISULFID        241      267
FT      DISULFID        274      316
FT      DISULFID        302      327
FT      DISULFID        336      378
FT      DISULFID        364      389
FT      DISULFID        396      439
FT      DISULFID        425      450
FT      DISULFID        454      505
FT      DISULFID        486      515
FT      DISULFID        524      567
FT      DISULFID        553      578
FT      DISULFID        582      636
FT      DISULFID        616      646
FT      CARBOHYD        162      162
FT      CARBOHYD        545      545
FT      SITE           617      619
FT      VARIANT         450      450
FT      SEQUENCE        661 AA; 75491 MW; 57A2FB46560857F2 CRC64;
Query Match
Best Local Similarity 24.0%; Score 764.5; DB 1; Length 661;
Matches 195; Conservative 85; Mismatches 254; Indels 139; Gaps 21;
QY      3  LLFSVILISWVTVGGEGTLCDFPKIHGHFLYDEEDYNPFPSQVPTGEV--FYYS----- 54
Db      6  LTFIILIL-----ISGE-----LYAEKPCGGFPVHVGRIAQYYTFKSPFYF 47
QY      55 -----CEYFNFSVPSKSFWRITCTCEGHSWPTPKCLRMCSFPFVKNGHSESSGLI 103
Db      48  PMSIDKKLSPFCLAGYTTESGRQEQTCTCEGWSPEPRCFKKCTKPDLSNGYISDVKLL 107
QY      104 H-----LEG----- 107
Db      108 YKIQENMHYGCASGYKTTGGKDDEVVCLDSQSSQPTCRKEHETCLAPELYLNGYNSTQ 167
QY      108 -----DTVQLICNTGYSLONNEK--NISCVERGWSTPPICSFTKGECVPILEANVDAQ 159

```

QY	241	FIINGPKKIQCVGDGENTTTLPTC	VEQVKTCGYIPELEYGYVQSPVPYQHGVSV	EVNCRNE	300
Db	152	-----	II:::II	-----	II:::II
QY	301	YAMIGNMITCINGIWT	ELPMCVATHOLKCKRCKIAGVNIKTLK	SGKENHNSRIRYCS	360
Db	179	YATADGN	-----	-----	-----
QY	361	DIFRYHVSVCINGKN	WPNVDCTEKREQFCPPPPQIPNAQNMTTV	YQDGEKVAVLCKEN	420
Db	186	-----	-----	-----	-----
QY	421	YLLPEAKEIVCKDGRW	QWSILPRCVSTAYCGPPPSINNGDTTSF	PLSVYPPGSTVTYRCQS	480
Db	186	-----	SSGSTITCQLONGWSAQPI	CINSEKCGPPPI	ISNGDTTSFLKLVVPOSRVYQCQP
QY	481	FYKLGSVTVTCRNKQ	WSEPPRCLOPCVVSSENMKNKIOLKWRNDG	LKYAKTGDAVEFQ	540
Db	242	YYELGVSNYVTCN	GEWSEPPRCIIPCIIITENMKNKI	LKGRSDRKYAKTGDTIEFM	301
QY	541	KXPHKAMITSSPPF	RAICQEGKFEPICE	569	
Db	302	CKLGYNANT	SILSFOAVCREGIVEYPRCE	330	
RESULT 9					
FHR4_HUMAN					
ID	FHR4_HUMAN	STANDARD;	PRT;	331 AA.	
AC	Q92496;	Q9UJY6;			
DT	16-OCT-2001	(Rel. 40, Created)			
DT	16-OCT-2001	(Rel. 40, Last sequence update)			
DT	15-JUN-2002	(Rel. 41, Last annotation update)			
DE	Complement factor H-related protein 4 precursor (FHR-4).				
GN	FHR4				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Liver;				
RR	MEDLINE=97190290; PubMed=9038172;				
RR	Skerka C., Hellwege J., Weber W., Tilkorn A., Buck F., Marti T.,				
RR	Kampen E., Beisiegel U., Zipfel P.F.;				
RR	"The human factor H-related protein 4 (FHR-4). A novel short consensus				
RR	repeat-containing protein is associated with human triglyceride-rich				
RR	lipoproteins.";				
RR	J. Biol. Chem. 272:5627-5634(1997).				
RR	[2]				
RR	SEQUENCE OF 1-19 FROM N.A.				
RR	Male D.A., Ormsby R.J., Giannakis E., Gordon D.L.;				
RR	"Promoter region of complement factor H-related 4 (fhr-4) gene.";				
RR	Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.				
RR	[3]				
RR	CHARACTERIZATION.				
RR	MEDLINE=98136603; PubMed=9476126;				
RR	Hellwege J., Skerka C., Zipfel P.F.;				
RR	"Biochemical and functional characterization of the factor-H-related				
RR	protein 4 (FHR-4)."				
RR	Immunopharmacology 38:149-157(1997).				
CC	-!- FUNCTION: INVOLVED IN COMPLEMENT REGULATION. CAN ASSOCIATE WITH				
CC	LIPOPROTEINS AND MAY PLAY A ROLE IN LIPID METABOLISM.				
CC	-!- SUBUNIT: HOMODIMER.				
CC	-!- SUBCELLULAR LOCATION: Extracellular.				
CC	-!- PTM: GLYCOSYLATED.				
CC	-!- SIMILARITY: CONTAINS 5 SUSHI (SCR) DOMAINS.				
CC	-!- SIMILARITY: STRONG, TO FACTOR H.				
CC	-----				
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```
CC -----
CC EMBL: X98337; CAA66980.1; -.
CC DR EMBL: AFI90816; AAF05951.1; -.
CC DR HSSP: P10998; 1WVD.
CC MIM: 605337; -.
CC InterPro: IPR000436; Sushi_SCR_CCP.
CC Pfam: PF00084; sushi; 5.
CC SMART: SM00032; CCP; 4.
KW Repeat; Glycoprotein; Sushi; Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 331 COMPLEMENT FACTOR H-RELATED PROTEIN 4.
FT DOMAIN 23 85 SUSHI 1.
FT DOMAIN 87 146 SUSHI 2.
FT DOMAIN 148 205 SUSHI 3.
FT DOMAIN 210 266 SUSHI 4.
FT DOMAIN 268 331 SUSHI 5.
FT DISULFID 24 73 BY SIMILARITY.
FT DISULFID 56 84 BY SIMILARITY.
FT DISULFID 88 134 BY SIMILARITY.
FT DISULFID 117 145 BY SIMILARITY.
FT DISULFID 149 193 BY SIMILARITY.
FT DISULFID 176 204 BY SIMILARITY.
FT DISULFID 211 254 BY SIMILARITY.
FT DISULFID 240 265 BY SIMILARITY.
FT DISULFID 269 320 BY SIMILARITY.
FT DISULFID 303 330 BY SIMILARITY.
FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 186 186 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 206 206 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 310 310 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 331 AA; 37325 MW; 5A0D04AB84B841424 CRC64;

Query Match 21.4%; Score 681; DB 1; Length 331;
Best Local Similarity 28.3%; Pred. No. 5e-43;
Matches 163; Conservative 48; Mismatches 114; Indels 250; Gaps 11;

QY 1 MLLLESVILSWTVGG-EGTLCDRPKIHGFLYDEEDYNPFQSVPTGVVYSCRYNF 59
DB 1 MLLINVLILWVSCANGQEVKCDPFEOIHGGLYKSLRLRFPAAAGOSYSYCDQNF 60
QY 60 VSPSKSFWRITTECGWSPPTKCLRMCSFP--FVKNHSESSGLIHLEGDVQIIICNTG 117
DB 61 VTPGSYWDYHCTQDGLWLPVCLRTCKSDIEIENGFISSSVILNKEIQYCKPG 120
QY 118 YSLQ--NNEKNISCVRGWSTPPIC-SFTTGECHVPILNVAQPKESYKVGDLVKFS 174
DB 121 YATADNGSGSITCLQNGWSAQPICIKF----CDMPVFE-NSRAKSNMRFKLHDTLDYE 175
QY 175 CRKNLIRVGSYSVQYQFGWSPNFPPTCKGQVRSGPPPLSQLSNGEVKEIRKEEYGHNEVVE 234
DB 176 C----- 176
QY 235 YDCNPNIINGPKKIQCVGDGWTTLPTCVQVKTGYIPELEXGYQVQPPYQHGVSVE 294
DB 177 YD-----GY--EISYG----- 185
QY 295 VNCREYAMIGNNMITGINGIWTLPVATHQLRKCKIAGVNIKTLKLSGREFNHSR 354
DB 186 ----- 185
QY 355 IRYRCSDFIRYHSCVINGKNWPEVDCTEKREQFCPPPPQIPNAQNMTTVNYODGEKVA 414
DB 186 -----NTGS----- 190
QY 415 VLCKENYLLPEAKIEICKGRWQSLPRCVSTAYCGPPPSINNGDTSFPLSVYPPGSTV 474
DB 191 -----IVCGEGWGHFFTCYNSSEKCGPPPLISNGDTSFLLKVVYPOSRV 236
QY 475 TYRQSFYKLGQSVTVTCRNKQWSEPPRCPLDPCVVSSEENMKNKNIQLKWRDGLKYARTG 534
DB 475 :||||:|||| ||| : |||||: ||:::||||||| ||| : |||||
```

```
Db 237 EYQCQSYELQGSNYVTCSNGWSEPPRCIHPCIITEENMKNKNIQLKGSIDIKYAKTG 296
QY 535 DAVEFOCKFPKHAMISSPPFRAICQCKGKFFVPC 569
DB 297 DTIEFMCKLGYNANTSVLSFQAVCREGIVETPRCE 331

RESULT 10
CR2_HUMAN
ID CR2_HUMAN STANDARD; PRT; 1033 AA.
AC P20023; Q13866; Q14212;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Complement receptor type 2 precursor (Cr2) (Complement C3d receptor)
DE (Epstein-Barr virus receptor) (EBV receptor) (CD21 antigen).
GN CR2 OR C3DR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89123277; PubMed=2563370;
RA Fujisaku A., Harley J.B., Frank M.B., Gruner B.A., Frazier B.,
RA Holers V.M.;
RT "Genomic organization and polymorphisms of the human C3d/Epstein-Barr
RT virus receptor";
RL J. Biol. Chem. 264:2118-2125(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=B-cell;
RP CR2_HUMAN;
RX MEDLINE=88171282; PubMed=2832506;
RA Weis J.J., Toothaker L.E., Smith J.A., Weis J.H., Fearon D.T.;
RT "Structure of the human B lymphocyte receptor for C3d and the
RT Epstein-Barr virus and relatedness to other members of the family
RT C3/C4 binding proteins.";
RL J. Exp. Med. 167:1047-1066(1988).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM C).
RX MEDLINE=88097454; PubMed=282171;
RA Moore M., Cooper N., Rack B., Nemerow G.;
RT "Molecular cloning of the cDNA encoding the Epstein-Barr virus C3d
RT receptor (complement receptor type 2) of human B lymphocytes.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:9194-9198(1987).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORMS A; C AND D).
RX MEDLINE=99165388; PubMed=10068037;
RA Barel M., Balbo M., Frade R.;
RT "Evidence for a new transcript of the Epstein-Barr virus/C3d receptor
RT (CR2, CD21) which is due to alternative exon usage.";
RL Mol. Immunol. 35:1025-1031(1998).
RN [5]
RP SEQUENCE OF 226-233; 256-267; 332-341; 567-677 AND 898-908.
RX MEDLINE=86287311; PubMed=3016712;
RA Weis J.J., Fearon D.T., Klickstein L.B., Wong W.W., Richards S.A.,
RA de Bruyn Kops A., Smith J.A., Weis J.H.;
RT "Identification of a partial cDNA clone for the C3d/Epstein-Barr
RT virus receptor of human B lymphocytes: homology with the receptor for
RT fragments C3b and C4b of the third and fourth components of
RT complement.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:5639-5643(1986).
RN [6]
RP SEQUENCE OF 492-556 FROM N.A. (ISOFORM B).
RX MEDLINE=93294286; PubMed=8390533;
RA Sinha S.K., Todd S.C., Hedrick J.A., Speiser C.L., Lambris J.D.,
RA Tsoukas C.D.;
RT "Characterization of the EBV/C3d receptor on the human Jurkat T cell
RT line: evidence for a novel transcript.";
RL J. Immunol. 150:5311-5320(1993).
CC - FUNCTION: Receptor for complement C3d and for the Epstein-Barr
CC virus on human B-cells and T-cells. Participates in B lymphocytes
CC activation.
```

FT	DISULFID	51	82	BY SIMILARITY.
FT	DISULFID	91	132	BY SIMILARITY.
FT	DISULFID	118	146	BY SIMILARITY.
FT	DISULFID	154	197	BY SIMILARITY.
FT	DISULFID	183	210	BY SIMILARITY.
FT	DISULFID	215	256	BY SIMILARITY.
FT	DISULFID	242	271	BY SIMILARITY.
FT	DISULFID	276	325	BY SIMILARITY.
FT	DISULFID	305	342	BY SIMILARITY.
FT	DISULFID	351	393	BY SIMILARITY.
FT	DISULFID	379	406	BY SIMILARITY.
FT	DISULFID	410	453	BY SIMILARITY.
FT	DISULFID	439	466	BY SIMILARITY.
FT	DISULFID	471	509	BY SIMILARITY.
FT	DISULFID	495	522	BY SIMILARITY.
FT	DISULFID	527	576	BY SIMILARITY.
FT	DISULFID	556	593	BY SIMILARITY.
FT	DISULFID	602	644	BY SIMILARITY.
FT	DISULFID	630	657	BY SIMILARITY.
FT	DISULFID	662	699	BY SIMILARITY.
FT	DISULFID	685	714	BY SIMILARITY.
FT	DISULFID	719	762	BY SIMILARITY.
FT	DISULFID	748	779	BY SIMILARITY.
FT	DISULFID	788	830	BY SIMILARITY.
FT	DISULFID	816	843	BY SIMILARITY.
FT	DISULFID	851	894	BY SIMILARITY.
FT	DISULFID	880	907	BY SIMILARITY.
FT	DISULFID	912	955	BY SIMILARITY.
FT	DISULFID	941	968	BY SIMILARITY.
FT	CARBOHYD	121	121	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	127	127	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	172	194	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	372	372	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	492	492	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	623	623	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	682	682	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	800	800	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	823	823	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	861	861	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	911	911	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSP LIC	499	534	MISSING (IN ISOFORM B).
FT	VARSP LIC	525	556	ITCPCPPVYNGAHGSSLEDFPYGTVVYTC -> NHLTPT
FT	VARSP LIC	659	659	TPCYLQWTHREFLRGSLEFWHNGHLHM (IN ISOFORM B).
FT	VARSP LIC	716	723	K -> KGQCPPLGLHGHRTGNTVFFVSGMTVDYTCDPGY
FT	VARSP LIC	457	457	LVLVGNCHCMPSGNWSPADPCE (IN ISOFORM C
FT	VARSP LIC	646	646	AND ISOFORM D).
FT	VARSP LIC	667	667	MISSING (IN ISOFORM D).
FT	VARSP LIC	759	787	MISSING (IN REF. 2).
FT	VARSP LIC	886	886	A -> R (IN REF. 3).
FT	VARSP LIC	890	890	Q -> D (IN REF. 5).
FT	VARSP LIC	902	902	KLQCRSDSKHGHSWSPQCLRSPPVTR -> NCSAEVIL
FT	VARSP LIC	906	906	KAWILERAPPOCLRLSL (IN REF. 3).
FT	VARSP LIC	993	993	L -> V (IN REF. 3).
FT	VARSP LIC	1003	1003	A -> P (IN REF. 3).
FT	VARSP LIC	1033	1033	Q -> G (IN REF. 5).
FT	VARSP LIC	112973	112973	H -> L (IN REF. 5).
FT	VARSP LIC	112973	112973	I -> V (IN REF. 2).
FT	VARSP LIC	112973	112973	E -> A (IN REF. 2).
FT	VARSP LIC	112973	112973	MM; 1749DBA407847ADA CRC64;
FT	VARSP LIC	112973	112973	Query Match 12.9%; Score 411; DB 1; Length 1033;
FT	VARSP LIC	112973	112973	Best Local Similarity 19.9%; Pred. No. 1.3e-22;
FT	VARSP LIC	112973	112973	Matches 169; Conservative 103; Mismatches 244; Indels 332; Gaps 40
QY	26	PKIHGFLYDEEDYNPFQVPTGEVFFYSCYENFVSPKSEFWTRITCTEG-WSPTP---	81	
Db	158	PMIHNGHHTSEN-----VGSIAPLGVSIVTSCESGYLLVGEKI-----INCLSSGKRSVAPPTC	210	
QY	82	---KCLRMCSPFPVKNGHSESSGLIHLEGGDTVQICMTGYSLQNNKNIISCVBERG-----W	134	
Db	211	BEARCKSLGRP---NGVKKEPPIILRV-GVTANFFCDEGRLQGPSS-FVTAGOGVAM	265	


```
FT DOMAIN 420 488 SUSHI A7.
FT DOMAIN 492 550 SUSHI B1.
FT DOMAIN 553 612 SUSHI B2.
FT DOMAIN 615 683 SUSHI B3.
FT DOMAIN 687 744 SUSHI B4.
FT DOMAIN 746 804 SUSHI B5.
FT DOMAIN 807 867 SUSHI B6.
FT DOMAIN 870 938 SUSHI B7.
FT DOMAIN 942 1000 SUSHI C1.
FT DOMAIN 1003 1062 SUSHI C2.
FT DOMAIN 1065 1133 SUSHI C3.
FT DOMAIN 1137 1194 SUSHI C4.
FT DOMAIN 1196 1254 SUSHI C5.
FT DOMAIN 1257 1317 SUSHI C6.
FT DOMAIN 1320 1388 SUSHI C7.
FT DOMAIN 1395 1453 SUSHI D1.
FT DOMAIN 1456 1515 SUSHI D2.
FT DOMAIN 1518 1586 SUSHI D3.
FT DOMAIN 1590 1647 SUSHI D4.
FT DOMAIN 1649 1707 SUSHI D5.
FT DOMAIN 1710 1770 SUSHI D6.
FT DOMAIN 1773 1841 SUSHI D7.
FT DOMAIN 1847 1905 SUSHI E1.
FT DOMAIN 1908 1966 SUSHI E2.
FT DISULFID 43 86 BY SIMILARITY.
FT DISULFID 73 99 BY SIMILARITY.
FT DISULFID 104 145 BY SIMILARITY.
FT DISULFID 131 161 BY SIMILARITY.
FT DISULFID 166 215 BY SIMILARITY.
FT DISULFID 215 232 BY SIMILARITY.
FT DISULFID 238 280 BY SIMILARITY.
FT DISULFID 266 293 BY SIMILARITY.
FT DISULFID 297 340 BY SIMILARITY.
FT DISULFID 326 353 BY SIMILARITY.
FT DISULFID 358 400 BY SIMILARITY.
FT DISULFID 386 416 BY SIMILARITY.
FT DISULFID 421 470 BY SIMILARITY.
FT DISULFID 450 487 BY SIMILARITY.
FT DISULFID 493 536 BY SIMILARITY.
FT DISULFID 523 549 BY SIMILARITY.
FT DISULFID 554 595 BY SIMILARITY.
FT DISULFID 581 611 BY SIMILARITY.
FT DISULFID 616 665 BY SIMILARITY.
FT DISULFID 645 682 BY SIMILARITY.
FT DISULFID 688 730 BY SIMILARITY.
FT DISULFID 716 743 BY SIMILARITY.
FT DISULFID 747 790 BY SIMILARITY.
FT DISULFID 776 803 BY SIMILARITY.
FT DISULFID 808 850 BY SIMILARITY.
FT DISULFID 836 866 BY SIMILARITY.
FT DISULFID 871 920 BY SIMILARITY.
FT DISULFID 900 937 BY SIMILARITY.
FT DISULFID 943 986 BY SIMILARITY.
FT DISULFID 973 999 BY SIMILARITY.
FT DISULFID 1004 1045 BY SIMILARITY.
FT DISULFID 1031 1061 BY SIMILARITY.
FT DISULFID 1066 1115 BY SIMILARITY.
FT DISULFID 1095 1132 BY SIMILARITY.
FT DISULFID 1138 1180 BY SIMILARITY.
FT DISULFID 1166 1193 BY SIMILARITY.
FT DISULFID 1197 1240 BY SIMILARITY.
FT DISULFID 1226 1253 BY SIMILARITY.
FT DISULFID 1258 1300 BY SIMILARITY.
FT DISULFID 1286 1316 BY SIMILARITY.
FT DISULFID 1321 1370 BY SIMILARITY.
FT DISULFID 1350 1387 BY SIMILARITY.
FT DISULFID 1396 1439 BY SIMILARITY.
FT DISULFID 1426 1452 BY SIMILARITY.
FT DISULFID 1457 1498 BY SIMILARITY.
FT DISULFID 1484 1514 BY SIMILARITY.
FT DISULFID 1519 1568 BY SIMILARITY.
FT DISULFID 1548 1585 BY SIMILARITY.
FT DISULFID 1591 1633 BY SIMILARITY.

Query Match 12.1%; Score 385; DB 1; Length 2039;
Best Local Similarity 25.4%; Pred No. 2.3e-20;
Matches 142; Conservative 78; Mismatches 204; Indels 136; Gaps 33;

QY 37 EDYNPFSQVPT----GEVFIYSCYNF---VSPSKSFWTTRITCEGWSP--PKLRMCS 88
DB 1462 EPFNGMVHINTDQFGSTVNYSCNEGFRILGPS-----TTCLVSGNNVTDWKAPICE 1515
QY 89 F-----PFVKNG--HSESSGLHLBGDTVOICTNGYSLON-----NEKNISCVBERG-- 133
DB 1516 IISCEPPPTISNGDFYSNNRTSFH--NGTVVYQCHTGPGEQLELFELGERSIYCTSKDDQ 1574
QY 134 ---WSTPPICSTKGECHVPILANVDQAQPKESYKVGVDVLFKSCRNLRVGSDSVQCY 190
DB 1575 VGVWSSPPRCISTNKCTAPEVENAIRVPCNRSFFSLTEIRFCQPGFVWGSHTVQCO 1634
QY 191 QFG--WSPNFTCKQVRSCGPPPOLSGNGEVKEIRKEEYGHNEVVEYDCNPFTIINGPKKI 249
DB 1635 TNGRWGPKLPHCS---RVCQPPPEILHGEHTLSHQDNFSPGQEVFVSCPEPSYDLRGAASL 1691
QY 250 QCV--DGEWT--TLPTCYEQVQKTC--GYIPELEYGVYQVSPV--PYOHGVSVEVNCNEVAMIG 305
DB 1692 HCTPQGDWSPEARCT--VRKCDDFLGQPLHGRV--LLPLNLQGLAKRVFVCDGEPRLKG 1747
QY 306 NNMITCINGIWTLPKMCVATHQKCKIAGVNIKTLLKLSGKKEFNHNSRIRYRSCSDFRY 365
DB 1748 RS-----ASH-----CVLAG--MKAL----- 1761
QY 366 RHVVCINGKNPEVDCTEKREOCPPPPOIPNNAQNMTT---TVNYODGEKVAVLCKEN-- 420
DB 1762 -----WNSVVPVCE--QIFCPNPPAILNGRHTGTPFGDIPY--GKEISVACDTHPD 1808
QY 421 -----YLLPEAKEIVCK-----DGRWQS--LPRCVEST--AYCGPPPSINNGDTTSFPISVVP 469
DB 1809 RGMTFNLIGESSIRCTSDPQGNVWSSPAPRCELSVPAACPHPKPQNGHYIGGHRVSLYL 1868
QY 470 PGSTVTRYCQSFYKLGQSVTVTCRNKO--WSE-----PPRCLDPCVVSEENKNKNIQK 522
DB 1869 PGMTISYTCDPGYLLVGKGFIFCTDQGIWSQLDHYCKEYVNCSPFLF-----MNGISKELE 1923
```

```

QY      523 WRNDGKLYAKTGDAVEFOCK 542
      : | : | | | : | :
Db      1924 MK---KVY-HYGDYVTLKCE 1939

```

```

RESULT 12
CR2_MOUSE
ID CR2_MOUSE STANDARD; PRT; 1025 AA.
AC P19070;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Complement receptor type 2 precursor (Cr2) (Complement C3d receptor).
GN CR2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=BALB/c;
RX MEDLINE=90229735; PubMed=2139457;
RA Fingerth J.D.;
RT "Comparative structure and evolution of murine CR2. The homolog of
RT the human C3d/EBV receptor (CD21).";
RL J. Immunol. 144:3458-3467(1990).
[2]
RN SEQUENCE OF 12-1025 FROM N.A.
RP MEDLINE=91010789; PubMed=2145366;
RX Molina H., Kinoshita T., Inoue K., Carel J.C., Holers V.M.;
RA "A molecular and immunochemical characterization of mouse CR2.
RA Evidence for a single gene model of mouse complement receptors 1 and
RT 2.";
RL J. Immunol. 145:2974-2983(1990).
[3]
RN SEQUENCE OF 343-401 AND 991-1025 FROM N.A.
RP MEDLINE=89098890; PubMed=2783485;
RX Fingerth J.D., Benedict M.A., Levy D.N., Strominger J.L.;
RA "Identification of murine complement receptor type 2.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:242-246(1989).
[4]
RN SEQUENCE OF 289-1025 FROM N.A.
RP MEDLINE=89381350; PubMed=2528587;
RX Kurtz C.B., Paul M.S., Asgester M., Weis J.J., Weis J.H.;
RA "Murine complement receptor gene family. II. Identification and
RT characterization of the murine homolog (Cr2) to human CR2 and its
RT molecular linkage to Cr1.";
RL J. Immunol. 143:2058-2067(1989).
CC -1- FUNCTION: RECEPTOR FOR COMPLEMENT C3D. PARTICIPATES IN B
CC LYMPHOCYTES ACTIVATION.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: B LYMPHOCYTES.
CC -1- SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION
CC (RCA) FAMILY.
CC -1- SIMILARITY: CONTAINS 15 SUSHI (SCR) DOMAINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collabor
CC between the Swiss Institute of Bioinformatics and the EMBL Outstat
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no
CC modified and this statement is not removed. Usage by and for comm
CC entities requires a license agreement (See http://www.lsb-sib.ch/anno
CC or send an email to license@lsb-sib.ch).
-----
DR EMBL; M81083; AAA37451.1; -
DR EMBL; M35684; AAA37448.1; -
DR EMBL; M61132; AAA63295.1; -
DR EMBL; M35685; AAA37450.1; ALT_SEQ.
DR EMBL; M29281; AAA37447.1; -
DR PIR; A43526; A43526.
DR HSP; P10996; IYVD.
DR MGD; MGI:88489; Cr2.
DR InterPro; IPR000436; Sushi_SCR_CCP.

```

```

FT CONFLICT 520 520 P -> A (IN REF. 2).
FT CONFLICT 962 963 MISSING (IN REF. 4).
SQ SEQUENCE 1025 AA; 112994 MW; 19E518B9A0273694 CRC64;

Query Match
  12.0%; Score 382; DB 1; Length 1025;
Best Local Similarity 24.2%; Pred. No. 1.7e-20;
Matches 158; Conservative 76; Mismatches 288; Indels 150; Gaps 36;

Qy 41 PFSQVP--TGEVFFYSC--EYNFVSPKSEFWTRITCTE--WSPTKCLR-----85
Db 26 PYSLPIVPGVNLRYTCSYRLIGERKAI-----CISENQVHATWDKAPICESVNTKI 80
Qy 86 MCSFPFVNGH-SESSGLIHLEGTVOILNTGYSLQNNKNSICV-----GWSPTPIC-140
Db 81 SCSDPIVPGFMNKGKAPFRHSDSVTFCTKANFTMKS-KTYMCQANEMWGTALPVE 139
Qy 141 SFTKGE-HPILRANVDAQPKKESYKVDVLFKSCRNLRVGSQVQYQFG-WSPNF 198
Db 140 SDFPLECPSLTIHNGHTGHVDQFVAGLSVYSCBPGYLLTGKTKIKLSSGDWDGVI 199
Qy 199 PTKGQVRSCGPPQLNGEYKEIRKEEYGHNEVVEYDCNPNFTINGPKKIQCVGDE---255
Db 200 PTKC--EAQCEHPGKFPNGQVKEPLSLQVG--TTVFESNEGVLQOGQPSQCVIVEOKA 255
Qy 256 -WTLPLTCVQVKTGCIPEL---EYGYVQPSVPPYQHGVSVEVNCRN-----EYAMI 304
Db 256 IWKPKVCKEIL--CPPPPVVRNGSHGTSFSENV-----YGSTVYTYTCDPSPEKGVSTLI 310
Qy 305 GNNMITCI-----NGIWT-ELPMCVATHQLKRC---KIAGVNIKTLLKLGKGFNHSRI 355
Db 311 GKTINCTGTSQKTIWNGPAPCVLSTSAVLCLQPKIKRGQILSLKDS---YSYNDTV 367
Qy 356 RYRCSDFIRVHSCVI-----NGKWNPEVDCERKQPCPPPPQIPNQAQNTT--TVNYQDG 410
Db 368 AFCEPGFTLKGNRISRCNAGHTWEPVPCVCEG--COAPPKIINGKEDSYLLNPDGP 424
Qy 411 EKVAVLCKENYLLPEAKEIYC-KDGRWQL-PRC-----442
Db 425 TSINYSQDGYLLVGEDTHCTPEGKWTPTPOCTVAECQVGPVPHLFRKPNQFIRAVN 484
Qy 443 -----VESTAY-----CGPPPSINNGDTSFPLSVYVPGSTV 474
Db 485 SSCDEGFQLESAYQLCQGTIPWFIEIRLCKEITCPPPVHNGTHWSSSEVDVPGTVV 544
Qy 475 TYRCOS-----FYKLOGSVTVTCRNK-----QWSEP-PRC--LDPCVSEENNNKNIQ 520
Db 545 TYMCYGPGERGVAKFLIGEOTICTSDSRGSGSWSPAPLCKLSLPAVQCTDVRHENGVR 604
Qy 521 LKWRNDGKLYAKTGDAVEFOCKFPHKAMISSPPFRAICQEGKP--EYPICE 569
Db 605 L---TDNKAPFYNDVSMFKC---DDGYILSGSSQIRCKANNTWDPEKPLCK 650

```

RESULT 13

```

C4BP_RAT
ID C4BP_RAT STANDARD; PRT; 558 AA.
AC Q63514;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE C4b-binding protein alpha chain precursor (C4bp).
GN C4BPA.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=101116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=97166082; PubMed=9013975;
RA Hillarp A., Wiklund H., Thern A., Dahlback B.;
RT "Molecular cloning of rat C4b binding protein alpha- and beta-chains:
structural and functional relationships among human, bovine, rabbit,

```

```

mouse, and rat proteins.";
J. Immunol. 158:1315-1323(1997).
-1- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT
ACTIVATION. IT BINDS AS A COPACTOR TO C3B/C4B INACTIVATOR
(C3BINA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT
ALSO ACCELERATES THE DEGRADATION OF THE C4BC2A COMPLEX (C3
CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA
CHAIN BINDS C4B. IT INTERACTS ALSO WITH ANTICOAGULANT PROTEIN S
AND WITH SERUM AMYLOID P COMPONENT.
-1- SUBUNIT: DISULFIDE-LINKED COMPLEX OF C4BP ALPHA AND BETA CHAINS.
-1- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.
-1- SIMILARITY: CONTAINS 8 SUSHI (SCR) DOMAINS.
-----
THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL: Z50051; CAA90391.1; -.
HESP: P10998; IYVD.
InterPro: IPR00436; SUSHI_SCR_CCP.
Pfam: PF00084; sush1; 8.
SMART: SM00032; CCP; 8.
KW Complement pathway; Plasma; Glycoprotein; Repeat; Sush1; Signal.
FT SIGNAL 1 13
FT CHAIN 14 558 C4B-BINDING PROTEIN ALPHA CHAIN.
FT DOMAIN 14 73 SUSHI 1.
FT DOMAIN 76 135 SUSHI 2.
FT DOMAIN 138 200 SUSHI 3.
FT DOMAIN 203 259 SUSHI 4.
FT DOMAIN 262 325 SUSHI 5.
FT DOMAIN 328 387 SUSHI 6.
FT DOMAIN 389 444 SUSHI 7.
FT DOMAIN 446 502 SUSHI 8.
FT DISULFID 15 60 BY SIMILARITY.
FT DISULFID 45 72 BY SIMILARITY.
FT DISULFID 77 118 BY SIMILARITY.
FT DISULFID 104 134 BY SIMILARITY.
FT DISULFID 139 182 BY SIMILARITY.
FT DISULFID 168 199 BY SIMILARITY.
FT DISULFID 204 246 BY SIMILARITY.
FT DISULFID 232 258 BY SIMILARITY.
FT DISULFID 263 312 BY SIMILARITY.
FT DISULFID 296 324 BY SIMILARITY.
FT DISULFID 328 351 BY SIMILARITY.
FT DISULFID 329 373 BY SIMILARITY.
FT DISULFID 336 386 BY SIMILARITY.
FT DISULFID 390 431 BY SIMILARITY.
FT DISULFID 417 443 BY SIMILARITY.
FT DISULFID 447 488 BY SIMILARITY.
FT DISULFID 474 501 BY SIMILARITY.
FT DISULFID 509 509 INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
FT DISULFID 521 521 INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
FT CARBOHYD 31 31 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 177 177 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 186 186 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 469 469 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 491 491 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 558 AA; 62266 MW; 592F0C667ED1E5FF CRC64;

```

```

Query Match
  10.7%; Score 339.5; DB 1; Length 558;
Best Local Similarity 24.9%; Pred. No. 1.1e-17;
Matches 126; Conservative 75; Mismatches 207; Indels 99; Gaps 30;

Qy 109 TVQIICNTGYSLQNNKNSICVVERG-WSTPPICSTFKGECHVP--ILEANVDAQPKKESY 165
Db 40 TLRYNCRPGYSRASSQSLYCKPLGKQWQINIA--VKKSCRNPGLQNGKVEV---KTDF 94
Qy 166 KVGVDVLKSCRNLRVGSDSVQCVQFG-----WSNPFPTCKGVRSQGGPPPLSGEYKE 221

```

Db 95 LFGSIEFSCSEGYILIGSTSYCEIOGKGVSWSDPLPEC--VIAKCGMPDINSNGKHG 152
 QY 222 IRKEEYGNVEVDCNENFIINGPKKIQ-----VDGEWTLTPCVCVEQVKTCGYIPELE 276
 Db 153 REEEFTYRSSVTKCDPDLFLLGNASTCTVNVKTVGWSPPPTCERI-ICPW-PKVL 210
 QY 277 YGYVQPSVP-PYOHGSVVEVNCNRYAMIGNMILTC-INGIWTLPMPVATHOLKRC-KI 333
 Db 211 HGTINSQFKHYKYKDSYVFCQKGVLRGSGVHCEADGWSPPVPC-----ELNSCTDI 266
 QY 334 AGVNIKTLLKSGKEFNNSRIR-----YRCSDDIPRHSVCGING----- 373
 Db 267 PDIPNAALIT-----SPRPRKEDYVPGTVLRY-----ICRPGYEPATROPMTVICOKD 315
 QY 374 -KNPEVDCTEKREQFCPPPPQIPNAQ-----NMFTTVYQDGEKVAVLCKENYLLPE 425
 Db 316 LSWMLRCKE-----ICCPVDPKRSVRVIOHEKAHPONDCTFFGDEVSYTCQNDIML-- 369
 QY 426 AKEIVCK-DGRWQ-SLPRCVSTAYCGPPPSINNGDTSFPLSVYVPGSTVYRQCSFYK 483
 Db 370 --TATCKSDGTWHPRTSPCHQS---CDFPPALAHGRTK---SSSYVVRTQTYECEGYR 422
 QY 484 LOGSVTVTCRNKQWS-EPRLCLDPC---VWSEENNNKNIQKWRNDGKLYAKTGDAVEF 539
 Db 423 LVGEATISQWYSQWTPAAPQCKALCKPEIGNVLSTNKDQ-----YVET-ENVTI 472
 QY 540 CKKEPHKAMISPPPPRAICQEGKFEY 566
 Db 473 QCDSCF-VMLGSQSI--TCSNGTWYP 496

RESULT 14

C4BP_BOVIN
 ID C4BP_BOVIN STANDARD; PRT; 610 AA.
 AC Q28065;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE C4B-binding protein alpha chain precursor (C4bp).
 GN C4BPA.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=95015909; PubMed=7930621;
 RT Hillarp A., Thern A., Dahlbaech B.;
 RT "Bovine C4b binding protein. Molecular cloning of the alpha- and
 RT beta-chains provides structural background for lack of complex
 RT formation with protein S".
 RL J. Immunol. 153:4190-4199(1994).
 CC -!- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT
 CC ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR
 CC (C3BI), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT
 CC ALSO ACCELERATES THE DEGRADATION OF THE C4B/C2A COMPLEX (C3
 CC CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A' ALPHA
 CC CHAIN BINDS C4B. IT INTERACTS ALSO WITH SERUM AMYLOID P COMPONENT.
 CC -!- SUBUNIT: DISULFIDE-LINKED COMPLEX OF C4BP ALPHA AND BETA CHAINS
 CC (BY SIMILARITY).
 CC -!- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.
 CC -!- SIMILARITY: CONTAINS 8 SUSHI (SCR) DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; Z31693; CAA83498.1; -.
 DR HSP: P10998; 1VVD.
 DR InterPro; IPR000436; Sushi_SCR_CCP.
 DR Pfam; PF00084; sushi; 8.
 DR SMART; SM00032; CCP; 8.
 KW Complement pathway; Plasma; Glycoprotein; Repeat; Sushi; Signal.
 FT SIGNAL 1 48
 FT CHAIN 49 610
 FT C4B-BINDING PROTEIN ALPHA CHAIN.
 FT DOMAIN 49 108
 FT SUSHI 1.
 FT DOMAIN 111 170
 FT SUSHI 2.
 FT DOMAIN 173 235
 FT SUSHI 3.
 FT DOMAIN 238 295
 FT SUSHI 4.
 FT DOMAIN 298 363
 FT SUSHI 5.
 FT DOMAIN 366 426
 FT SUSHI 6.
 FT DOMAIN 428 484
 FT SUSHI 7.
 FT DOMAIN 486 542
 FT SUSHI 8.
 FT BY SIMILARITY.
 FT DISULFID 50 95
 FT BY SIMILARITY.
 FT DISULFID 80 107
 FT BY SIMILARITY.
 FT DISULFID 112 153
 FT BY SIMILARITY.
 FT DISULFID 139 169
 FT BY SIMILARITY.
 FT DISULFID 174 217
 FT BY SIMILARITY.
 FT DISULFID 203 234
 FT BY SIMILARITY.
 FT DISULFID 239 281
 FT BY SIMILARITY.
 FT DISULFID 267 294
 FT BY SIMILARITY.
 FT DISULFID 299 350
 FT BY SIMILARITY.
 FT DISULFID 334 362
 FT BY SIMILARITY.
 FT DISULFID 366 390
 FT BY SIMILARITY.
 FT DISULFID 367 412
 FT BY SIMILARITY.
 FT DISULFID 429 471
 FT BY SIMILARITY.
 FT DISULFID 457 483
 FT BY SIMILARITY.
 FT DISULFID 487 528
 FT BY SIMILARITY.
 FT DISULFID 514 541
 FT BY SIMILARITY.
 FT DISULFID 549 549
 FT BY SIMILARITY.
 FT DISULFID 561 561
 FT CARBOHYD 66 66
 FT CARBOHYD 221 221
 FT CARBOHYD 525 525
 FT CARBOHYD 602 602
 FT SEQUENCE 610 AA; 68886 MW; D806B2708BA06B58 CRC64;
 Query Match 10.3%; Score 328; DB 1; Length 610;
 Best Local Similarity 22.8%; Pred. No. 8.7e-17;
 Matches 117; Conservative 76; Mismatches 218; Indels 102; Gaps 26;
 QY 107 GDTYQILICNTGYSLONNEKNISC-VERGWSPTTICSTFKGECVHP--ILEAVDQAPKE 163
 Db 73 GTTLRYTCRPGYRISRRKNFLICDGTDNWKYKEFC--YKKRCENFPELLNGQVIV---KT 127
 QY 164 SYKVGDLVKFSCRKNLIRVSDSVQY----QFGWSPNFPCTCKGVRSCTGPPQLSNGEV 219
 Db 128 DYSFGSEIEFSCSEGVVLIGSANSYCOLQDKGVVWSDPLQPC--IIAKCEPPTISNGRH 185
 QY 220 KEIRKEEYGHNEVEYDCNPNFIINGPKKIQ-----VDGEWTLTPCVCVEQVKTCGYIPE 274
 Db 186 NGGDEDFYTGSSVYSCDRDFSMGLKASISCRVENKKTIGVMSPPSPCKV-IC----- 239
 QY 275 LEYGVVQPSV-----PPYQHGVSVEVNCREYAMIGNMILTC-INGIWTLPMPCV 323
 Db 240 -----VQPVWDGKITSGFGPIYTYQQSIVIVACNKGFRLEGDSLHCADNSWNPDP 291
 QY 324 ATHQLKRCIAGVNIKTLLKLSGKEFNHNSRIRYRCSDI-----FRYRHS----- 368
 Db 292 PTCELNGC---LGLPHIPHALWERYDHTQTEQQQVYDIGFVLSYKCHGFKPETDGPPT 347
 QY 369 -VC-INGKWNPEVDCTEREQFCPPPPQIPNQNMT-----TTVNVQDEKAVLCK 418
 Db 348 VTCQNLWSPYIECKE-----VCCPEPNLNYGTSITLHRRPSTSTHCTYISGDKISYEC 403
 QY 419 ENYLLPEAKEIVC-KDGRWQSLPRCVSTAYCGPPPSINNGD---TTSFPLSVYPPGSTV 474
 Db 404 SKYMF-----DALCTKHGTWS---PRTPECRDCKSPFVIAHGOHKVYSKF-----FTDHOA 453

QY 475 TYRQSPYKLOGSVTVTCRNKQKSEP-PRCLDPVSEENWKNKNIQKWRNDOKLYAKT 533
 Db 454 VYEDKGYLVKAKELSTSGWSPVPOCKALCKPKEIYGRLSVE-----KVRYVE 506
 QY 534 GDAVEFOCKPFPKAMISPPFRATCQGGKPEYP 566
 Db 507 PEITIOCESGYVVGSE---NITCSEDRTWP 536

RESULT 15
 ID C4BP_HUMAN STANDARD; PRT; 597 AA.
 AC P04003;
 DT 23-OCT-1986 (Rel. 02, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE C4b-binding protein alpha chain precursor (C4bp) (Proline-rich
 DE protein) (PRP).
 GN CABPA OR C4BP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=90073699; PubMed=2590215;
 RA Matsuuchi T., Okamura S., Aso T., Sata T., Niho Y.;
 RT "Molecular cloning of the cDNA coding for proline-rich protein (PRP):
 RT identity of PRP as C4b-binding protein.";
 RL Biochem. Biophys. Res. Commun. 165:138-144(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91113199; PubMed=1989602;
 RA Aso T., Okamura S., Matsuuchi T., Sakamoto N., Sata T., Niho Y.;
 RT "Genomic organization of the alpha chain of the human C4b-binding
 RT protein gene.";
 RL Biochem. Biophys. Res. Commun. 174:222-227(1991).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RX Strausberg R.;
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 9-81 FROM N.A.
 RX MEDLINE=88242821; PubMed=3378624;
 RA Lintin S.J., Lewin A.R., Reid K.B.M.;
 RT "Derivation of the sequence of the signal peptide in human
 RT C4b-binding protein and interspecies cross-hybridisation of the C4bp
 RT cDNA sequence.";
 RL FEBS Lett. 232:328-332(1988).
 RN [5]
 RP SEQUENCE OF 203-288 FROM N.A.
 RX MEDLINE=86301119; PubMed=3017751;
 RA Lintin S.J., Reid K.B.M.;
 RT "Studies on the structure of the human C4b-binding protein gene.";
 RL FEBS Lett. 204:77-81(1986).
 RN [6]
 RP SEQUENCE OF 80-597 FROM N.A.
 RX MEDLINE=86025405; PubMed=3840370;
 RA Chung L.P., Bentley D.R., Reid K.B.M.;
 RT "Molecular cloning and characterization of the cDNA coding for C4b-
 RT binding protein, a regulatory protein of the classical pathway of the
 RT human complement system.";
 RL Biochem. J. 230:133-141(1985).
 RN [7]
 RP SEQUENCE OF 49-88.
 RX MEDLINE=85296001; PubMed=4033666;
 RA Chung L.P., Gagnon J., Reid K.B.M.;
 RT "Amino acid sequence studies of human C4b-binding protein: N-terminal
 RT proteolysis and alignment of the fragments produced by limited
 RT proteolysis with chymotrypsin and the peptides produced by cyanogen
 RT bromide treatment.";

RL Mol. Immunol. 22:427-435(1985).
 RN [8]
 RP ELECTRON MICROSCOPY, 3D-STRUCTURE, AND LIGAND BINDING.
 RX MEDLINE=83221615; PubMed=6222381;
 RA Dahlback B., Smith C.A., Mueller-Eberhard H.J.;
 RT "Visualization of human C4b-binding protein and its complexes with
 RT vitamin K-dependent protein S and complement protein C4b.";
 RL Proc. Natl. Acad. Sci. U.S.A. 80:3461-3465(1983).
 CC -1- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT
 CC ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR
 CC (C3BINA), WHICH THEN HYDROLYSES THE COMPLEMENT FRAGMENT C4b. IT
 CC ALSO ACCELERATES THE DEGRADATION OF THE C4bC2a COMPLEX (C3
 CC CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2a. ALPHA
 CC CHAIN BINDS C4b. IT INTERACTS ALSO WITH ANTICOAGULANT PROTEIN S
 CC AND WITH SERUM AMYLOID P COMPONENT.
 CC -1- SUBUNIT: DISULFIDE-LINKED COMPLEX OF C4BP ALPHA AND BETA CHAINS
 CC OF 3 POSSIBLE SORTS: A 570 kDa COMPLEX OF 7 ALPHA CHAINS AND 1
 CC BETA CHAIN, A 530 kDa HOMOPOLYMER OF ALPHA CHAINS OR A 500 kDa
 CC COMPLEX OF 6 ALPHA CHAINS AND 1 BETA CHAIN. THE CENTRAL BODY OF
 CC THE ALPHA CHAIN HOMOPOLYMER SUPPORTS TENTACLES, EACH WITH THE
 CC BINDING SITE FOR C4b AT THE END.
 CC -1- TISSUE SPECIFICITY: CHYLOMICRONS IN THE PLASMA.
 CC -1- SIMILARITY: CONTAINS 8 SUSHI (SCR) DOMAINS.
 CC -1- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.
 CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-17 IS THE INITIATOR.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M31452; AAA36507.1; -;
 DR EMBL; M62486; AAA36506.1; -;
 DR EMBL; M62475; AAA36506.1; JOINED.
 DR EMBL; M62476; AAA36506.1; JOINED.
 DR EMBL; M62477; AAA36506.1; JOINED.
 DR EMBL; M62478; AAA36506.1; JOINED.
 DR EMBL; M62479; AAA36506.1; JOINED.
 DR EMBL; M62480; AAA36506.1; JOINED.
 DR EMBL; M62481; AAA36506.1; JOINED.
 DR EMBL; M62482; AAA36506.1; JOINED.
 DR EMBL; M62484; AAA36506.1; JOINED.
 DR EMBL; M62485; AAA36506.1; JOINED.
 DR EMBL; BC022312; AAH22312.1; -;
 DR EMBL; X07853; CAA30701.1; -;
 DR EMBL; X04284; CAB51244.1; -;
 DR EMBL; X04296; CAA27839.1; -;
 DR EMBL; X02865; CAA26617.1; -;
 DR PIR; A33568; NBHUC4.
 DR HSSP; P10998; 1VVD.
 DR Genew; HGNC:1325; C4BPA.
 DR MIM; 120830; -;
 DR InterPro; IPR000436; Sushi_SCR_CCP.
 DR Pfam; PF00084; sushi; 8.
 DR SMART; SM00032; CCP; 8.
 DR Complement pathway; Plasma; Glycoprotein; Repeat; Sushi; Signal;
 KW Polymorphism. 1 48
 FT SIGNAL 49 597 C4b-BINDING PROTEIN ALPHA CHAIN.
 FT CHAIN 49 109 SUSHI 1.
 FT DOMAIN 112 171 SUSHI 2.
 FT DOMAIN 174 235 SUSHI 3.
 FT DOMAIN 238 295 SUSHI 4.
 FT DOMAIN 298 361 SUSHI 5.
 FT DOMAIN 364 423 SUSHI 6.
 FT DOMAIN 425 481 SUSHI 7.
 FT DOMAIN 483 539 SUSHI 8.
 FT DISULFID 50 96 BY SIMILARITY.
 FT DISULFID 81 108 BY SIMILARITY.
 FT DISULFID 113 154 BY SIMILARITY.

FT	DISULFID	140	170	BY SIMILARITY.
FT	DISULFID	175	217	BY SIMILARITY.
FT	DISULFID	203	234	BY SIMILARITY.
FT	DISULFID	239	281	BY SIMILARITY.
FT	DISULFID	267	294	BY SIMILARITY.
FT	DISULFID	299	348	BY SIMILARITY.
FT	DISULFID	332	360	BY SIMILARITY.
FT	DISULFID	364	7387	BY SIMILARITY.
FT	DISULFID	7365	7409	BY SIMILARITY.
FT	DISULFID	7399	422	BY SIMILARITY.
FT	DISULFID	426	468	BY SIMILARITY.
FT	DISULFID	454	480	BY SIMILARITY.
FT	DISULFID	484	525	BY SIMILARITY.
FT	DISULFID	511	538	BY SIMILARITY.
FT	DISULFID	546	546	INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
FT	DISULFID	558	558	INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
FT	CARBOHYD	221	221	N-LINKED (GLCNAC. . .).
FT	CARBOHYD	506	506	N-LINKED (GLCNAC. . .).
FT	CARBOHYD	528	528	N-LINKED (GLCNAC. . .).
FT	VARIANT	92	92	Q -> T.
FT	VARIANT	357	357	/FTID=VAR_001977.
FT	VARIANT	473	473	Y -> H.
FT	VARIANT	597	597	/FTID=VAR_001978.
FT	SEQUENCE	597	597	W -> L (IN DBSNP:1801341).
FT	SEQUENCE	597	597	/FTID=VAR_012038.
FT	SEQUENCE	597	597	67E03F2EA85A16DD CRC64;
QY	Query Match	9.5%;	Score 303;	DB 1; Length 597;
Db	Best Local Similarity	23.3%;	Pred. No. 5.8e-15;	
Db	Matches 104;	Conservative % 68;	Mismatches 194;	Indels 80; Gaps 26;
QY	107	GDATVQIICNTGYSLONNEKINISCVERG-WSTPPICSTFKGECHVP--ILLEANVDAQPKKE	163	
Db	74	GTTLTKYCLPGYVRSHSTQTLTNSDGEWYNTFCIIYKR--CRHPGELRNGQVEI---KT	128	
QY	164	SYKVGVDLKFSCRKNLIRVGSDSVQC---YQFGWSPNFPCTCKQVRSCGPPPOLSNGEV	219	
Db	129	DLSEFGQIEFSCSEGPFLLIGTTSRCEVQDRGVGWSHPLQOC--EIVKCRPPPDIRNGR-	185	
QY	220	KETRKEEYGHNEVVEYDCNPNFINGPKKIQOC-----VDGEWTTLPCTCBEQVKTGCIPE	274	
Db	186	HSGEENFYAGFYSVYSCDPFRSLLGHASISCTVENETIGVWRSPPTCEKI--TCRK-PD	243	
QY	275	LEYG-VYQPSVPYPQHGVSVEVNCREYAMIGNNMITC-INGIWTELPWCVATHQLRKCK	332	
Db	244	VSHGEWVGSGPIYNYKDTIVFKCKQGFVLRGSSVIHCDADSKWNPSPACEPNS---C-	299	
QY	333	IAGVNIKTLKLSGKEFNHNSR-----IRYRCSDIPIRYRHS-----VC-INGKNWP	377	
Db	300	---INLPDIPHASMETYPRTKEDVYVVGTVLRYRCHPGYKPTTDEPTTVICQKNLRWTP	356	
QY	378	EVDCTERREQFCPPPIQNAQ---NMTTVN---YQDEKVAVLCKENYLLPEAKEIV	430	
Db	357	YQGC-----EALCCPEPKLNNGEITQHRKSRPANHCVFYGDSEIFSCHET-----SRFSAI	408	
QY	431	CK-DGRWQSILPRCVESTAYGCPGPSINNGDTTSPPLSVYPPGST-----VTYRCQSFY	482	
Db	409	COGDGTSW--PRTPSCGDICNFPKPIAHGH-----YKQSSSYSPFKEELIIECDKGY	458	
QY	483	KLOGSVTVTCRNKOWSEP-PRCLDPC	507	
Db	459	ILVGQAKLSCSYSHWSAPAPQCKALC	484	

Search completed: July 17, 2003, 13:51:26
Job time : 29 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2003, 13:49:34 : Search time 43 Seconds
(without alignments)
1272.104 Million cell updates/sec

Title: US-09-805-337A-2

Perfect score: 3183

Sequence: 1 MLLLFVILISWSTVGEG.....SSPPFRAICQKGKFEYPICE 569

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_73:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1538	48.3	1231	1 NBHUH	complement factor
2	1524	47.9	808	2 D35069	complement factor
3	1427.5	44.8	1234	1 NBMSH	complement factor
4	1050.5	33.0	330	2 I56100	complement factor
5	927.5	29.1	270	2 I37278	complement factor
6	906.5	28.5	343	2 G35070	apolipoprotein H-r
7	861	27.0	669	2 S65551	factor H - bovine
8	783	24.6	668	2 A46013	coagulation factor
9	764.5	24.0	661	1 KFRU13	coagulation factor
10	710	22.3	331	2 A45222	complement factor
11	646	20.3	452	2 A35068	complement factor
12	476	15.0	1053	2 S46199	probable complemen
13	430	13.5	1091	1 PL0009	complement C3d/Eps
14	387.5	12.2	303	2 H35068	apolipoprotein H-r
15	387.5	12.2	2014	2 I36936	complement recepto
16	383	12.0	2489	2 I73012	complement C3b/C4b
17	382	12.0	676	2 A45900	complement C3b rec
18	382	12.0	1025	1 A43526	complement C3d/Eps
19	341.5	10.7	597	1 S53711	C4BP alpha chain p
20	340.5	10.7	579	2 A56740	sperm-egg recognit
21	339.5	10.7	558	2 S57953	C4BP protein alpha
22	328	10.3	610	1 I46001	C4b-binding protei
23	309.5	9.7	449	1 NBHUHS	complement factor
24	303	9.5	597	1 NBHUC4	C4b-binding protei
25	300.5	9.4	469	1 NBMSC4	C4b-binding protei
26	286.5	9.0	1797	2 T21889	hypothetical prote
27	286.5	9.0	1805	2 T21888	hypothetical prote
28	282.5	8.9	363	2 B45900	complement C3d/Eps
29	281	8.8	830	2 A30359	p-selectin precurs

ALIGNMENTS

RESULT 1 NBHUH

Complement factor H precursor, long splice form [validated] - human
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 08-Dec-2000
C:Accession: S00254; A60238; A54726; A61565; A26505; I72654; S66298
R:Ripoche, J.; Day, A.J.; Harris, T.J.R.; Sim, R.B.
Biochem. J. 249, 593-602, 1988

A:Title: The complete amino acid sequence of human complement factor H.
A:Reference number: S00254; MUID:88134059; PMID:2963625
A:Accession: S00254

A:Molecule type: mRNA

A:Residues: 1-1231 <RIP>

A:Cross-references: EMBL:Y00716; NID:g31964; PIDN:CAA68704.1; PID:g31965

A:Note: 402-Tyr was also found

A:Note: parts of this sequence, including the amino and carboxyl ends of the mature p

R:Estallor, C.; Schwaebel, W.; Pierich, M.; Weiss, E.H.

Eur. J. Immunol. 21, 799-802, 1991

A:Title: Human complement factor H: two factor H proteins are derived from alternativ

A:Reference number: A60238; MUID:91184292; PMID:1826264

A:Accession: A60238

A>Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-56;1177-1231 <EST>

A:Note: only portions of this 4.3 kilobase mRNA were sequenced

R:Day, A.J.; Ripoche, J.; Lyons, A.; McIntosh, B.; Harris, T.J.R.; Sim, R.B.

Biosci. Rep. 7, 201-207, 1987

A:Title: Sequence analysis of a cDNA clone encoding the C-terminal end of human compl

A:Reference number: A54726; MUID:88025472; PMID:2889480

A:Accession: A54726

A>Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 'DFRN',579-1231 <DAY>

A:Cross-references: GB:M17517; NID:g180497; PIDN:AAA52016.1; PID:g180498

A:Note: parts of this sequence were determined by protein sequencing

R:Ripoche, J.; Day, A.J.; Willis, A.C.; Belt, K.T.; Campbell, R.D.; Sim, R.B.

Biosci. Rep. 6, 65-72, 1986

A:Title: Partial characterization of human complement factor H by protein and cDNA se

A:Reference number: A61565; MUID:86188123; PMID:2938641

A:Accession: A61565

A>Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 'MEYGRNHLNNAKI',1050-1057,'T',1059-1102 <RI2>

R:Sim, R.B.; Disclipio, R.G.

Biochem. J. 205, 285-293, 1982

A:Title: Purification and structural studies on the complement-system control protein

A:Reference number: A26505; MUID:83048213; PMID:6215918

A:Accession: A26505

A:Molecule type: protein

A:Residues: 19-20,'Q',22-29,'V',31-33,'Q',35 <SIM>

R:Barlow, P.N.; Norman, D.G.; Steinkasserer, A.; Horne, T.J.; Pearce, J.; Driscoll, P.

Biochemistry 31, 3626-3634, 1992

complement regulat
E-selectin precurs
scavenger receptor
hypothetical prote
E-selectin precurs
complement C3b/C4b
P-selectin precurs
P-selectin - rat
P-selectin precurs
complement control
endothelial leukoc
apolipoprotein H p
complement recepto
seizure-related pr
apolipoprotein H p
apolipoprotein H p

30 276.5 8.7 497 2 JC2054
31 274.5 8.6 612 2 B42755
32 266.5 8.4 2043 2 T18524
33 264.5 8.3 560 2 T16833
34 264.5 8.3 610 2 A35046
35 255.5 8.0 482 2 A34924
36 255 8.0 768 2 A42755
37 253 7.9 768 2 I53821
38 247.5 7.8 646 2 JN0473
39 246 7.7 360 2 TA2921
40 244.5 7.7 551 2 I46709
41 243 7.6 345 1 NBMS
42 240 7.5 440 2 A43519
43 238 7.5 977 2 I52657
44 237.5 7.5 345 1 JN0465
45 234 7.4 345 1 NBHU

RESULT 2
D35069
complemen
C; Species


```

F:936-989/Domain: complement factor H repeat homology <FH16>
F:994-1048/Domain: complement factor H repeat homology <FH17>
F:1053-1107/Domain: complement factor H repeat homology <FH18>
F:1114-1168/Domain: complement factor H repeat homology <FH19>
F:1172-1233/Domain: complement factor H repeat homology <FH20>
F:21-66,52-80,85-129,114-141,146-192,178-205,210-251,237-262,267-309,294-320,325-374,357-377,381-400,404-420,424-440,444-460,464-480,484-500,504-520,524-540,544-560,564-580,584-600,604-620,624-640,644-660,664-680,684-700,704-720,724-740,744-760,764-780,784-800,804-820,824-840,844-860,864-880,884-900,904-920,924-940,944-960,964-980,984-1000,1004-1020,1024-1040,1044-1060,1064-1080,1084-1100,1104-1120,1124-1140,1144-1160,1164-1180,1184-1200,1204-1220,1224-1240,1244-1260,1264-1280,1284-1300,1304-1320,1324-1340,1344-1360,1364-1380,1384-1400,1404-1420,1424-1440,1444-1460,1464-1480,1484-1500,1504-1520,1524-1540,1544-1560,1564-1580,1584-1600,1604-1620,1624-1640,1644-1660,1664-1680,1684-1700,1704-1720,1724-1740,1744-1760,1764-1780,1784-1800,1804-1820,1824-1840,1844-1860,1864-1880,1884-1900,1904-1920,1924-1940,1944-1960,1964-1980,1984-2000,2004-2020,2024-2040,2044-2060,2064-2080,2084-2100,2104-2120,2124-2140,2144-2160,2164-2180,2184-2200,2204-2220,2224-2240,2244-2260,2264-2280,2284-2300,2304-2320,2324-2340,2344-2360,2364-2380,2384-2400,2404-2420,2424-2440,2444-2460,2464-2480,2484-2500,2504-2520,2524-2540,2544-2560,2564-2580,2584-2600,2604-2620,2624-2640,2644-2660,2664-2680,2684-2700,2704-2720,2724-2740,2744-2760,2764-2780,2784-2800,2804-2820,2824-2840,2844-2860,2864-2880,2884-2900,2904-2920,2924-2940,2944-2960,2964-2980,2984-3000,3004-3020,3024-3040,3044-3060,3064-3080,3084-3100,3104-3120,3124-3140,3144-3160,3164-3180,3184-3200,3204-3220,3224-3240,3244-3260,3264-3280,3284-3300,3304-3320,3324-3340,3344-3360,3364-3380,3384-3400,3404-3420,3424-3440,3444-3460,3464-3480,3484-3500,3504-3520,3524-3540,3544-3560,3564-3580,3584-3600,3604-3620,3624-3640,3644-3660,3664-3680,3684-3700,3704-3720,3724-3740,3744-3760,3764-3780,3784-3800,3804-3820,3824-3840,3844-3860,3864-3880,3884-3900,3904-3920,3924-3940,3944-3960,3964-3980,3984-4000,4004-4020,4024-4040,4044-4060,4064-4080,4084-4100,4104-4120,4124-4140,4144-4160,4164-4180,4184-4200,4204-4220,4224-4240,4244-4260,4264-4280,4284-4300,4304-4320,4324-4340,4344-4360,4364-4380,4384-4400,4404-4420,444-4460,4464-4480,4484-4500,4504-4520,4524-4540,4544-4560,4564-4580,4584-4600,4604-4620,4624-4640,4644-4660,4664-4680,4684-4700,4704-4720,4724-4740,4744-4760,4764-4780,4784-4800,4804-4820,4824-4840,4844-4860,4864-4880,4884-4900,4904-4920,4924-4940,4944-4960,4964-4980,4984-5000,5004-5020,5024-5040,5044-5060,5064-5080,5084-5100,5104-5120,5124-5140,5144-5160,5164-5180,5184-5200,5204-5220,5224-5240,5244-5260,5264-5280,5284-5300,5304-5320,5324-5340,5344-5360,5364-5380,5384-5400,5404-5420,5424-5440,5444-5460,5464-5480,5484-5500,5504-5520,5524-5540,5544-5560,5564-5580,5584-5600,5604-5620,5624-5640,5644-5660,5664-5680,5684-5700,5704-5720,5724-5740,5744-5760,5764-5780,5784-5800,5804-5820,5824-5840,5844-5860,5864-5880,5884-5900,5904-5920,5924-5940,5944-5960,5964-5980,5984-6000,6004-6020,6024-6040,6044-6060,6064-6080,6084-6100,6104-6120,6124-6140,6144-6160,6164-6180,6184-6200,6204-6220,6224-6240,6244-6260,6264-6280,6284-6300,6304-6320,6324-6340,6344-6360,6364-6380,6384-6400,6404-6420,6424-6440,6444-6460,6464-6480,6484-6500,6504-6520,6524-6540,6544-6560,6564-6580,6584-6600,6604-6620,6624-6640,6644-6660,6664-6680,6684-6700,6704-6720,6724-6740,6744-6760,6764-6780,6784-6800,6804-6820,6824-6840,6844-6860,6864-6880,6884-6900,6904-6920,6924-6940,6944-6960,6964-6980,6984-7000,7004-7020,7024-7040,7044-7060,7064-7080,7084-7100,7104-7120,7124-7140,7144-7160,7164-7180,7184-7200,7204-7220,7224-7240,7244-7260,7264-7280,7284-7300,7304-7320,7324-7340,7344-7360,7364-7380,7384-7400,7404-7420,7424-7440,7444-7460,7464-7480,7484-7500,7504-7520,7524-7540,7544-7560,7564-7580,7584-7600,7604-7620,7624-7640,7644-7660,7664-7680,7684-7700,7704-7720,7724-7740,7744-7760,7764-7780,7784-7800,7804-7820,7824-7840,7844-7860,7864-7880,7884-7900,7904-7920,7924-7940,7944-7960,7964-7980,7984-8000,8004-8020,8024-8040,8044-8060,8064-8080,8084-8100,8104-8120,8124-8140,8144-8160,8164-8180,8184-8200,8204-8220,8224-8240,
```

```
Db 180 YEMFDEEVMCLNGNWT----- 197
QY 361 DIFRYRHSVCINGKWNPEVDCTEKREQFCPPPPQIPNAQNMNTTVNYQDGEKVAVLCKEN 420
Db 198 -----PPQ----- 200
QY 421 YLLPEAKEIVCKDGRWQSLPRCVSTAYCGPPPSINNGDTSFPLSVYPPGSTVTVYRQCS 480
Db 201 -----CKD-----STGKCGPPPPIDNGDITSFPLSVYAPASSVEYQCCN 239
QY 481 FYKLOGSVTVTCRNKQWSEPPRCPLDPCVVSENNMKNKNIQLKWRNDGKLYAKTGDAVEFO 540
Db 240 LYQLEGNRIITCRNQWSEPPKCLHPCVISREIMENYIALRWTAOKOLYLRTGESAEFV 299
QY 541 CKFPHKAMISSPPFRAICQEGKFEYPIC 568
Db 300 CKRGYRLSSRSHTLTTCWDGKLEYPIC 327

RESULT 5
I37278
complement factor H-related protein 2 precursor - human
N:Alternate names: FHR-2
C:Species: Homo sapiens (man)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 21-Jul-2000
C:Accession: I37278; I37388; S24564
R:Skerka, C.; Timmann, C.; Horstmann, R.D.; Zipfel, P.F.
J. Immunol. 148, 3313-3318, 1992
A:Title: Two additional human serum proteins structurally related to complement factor H
A:Reference number: I37278; MUID:92251200; PMID:1533657
A:Accession: I37278
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-270 <ZIP>
A:Cross-references: EMBL:X64877; NID:g30498; PIDN:CAA46096.1; PID:g30499
R:Skerka, C.; Moulds, J.M.; Tailon-Miller, P.; Hourcade, D.; Zipfel, P.F.
Immunogenetics 42, 268-274, 1995
A:Title: The human factor H-related gene 2 (FHR2): structure and linkage to the coagulation
A:Reference number: I37388; MUID:95402981; PMID:7672821
A:Accession: I37388
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-84, 'K', 86-143, 'S', 172-270 <SKE>
A:Cross-references: EMBL:X86564; NID:g1064907; PIDN:CAA60375.1; PID:g1064908
C:Genetics:
A:Gene: GDB:HFL3; FHR2
A:Cross-references: GDB:132658
A:Map position: Iq32-Iq32
A:Introns: 20/1; 85/1; 144/1; 171/1; 205/1
C:Superfamily: apolipoprotein H; complement factor H repeat homology
C:Keywords: duplication; extracellular protein; glycoprotein; tandem repeat
F:23-83/Domain: complement factor H repeat homology <FHL>
F:87-140/Domain: complement factor H repeat homology <FHR2>
F:149-203/Domain: complement factor H repeat homology <FHR3>
F:207-267/Domain: complement factor H repeat homology <FHR4>
F:126/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 29.1%; Score 927.5; DB 2; Length 270;
Best Local Similarity 34.6%; Pred. No. 2.5e-56;
Matches 198; Conservative 20; Mismatches 47; Indels 307; Gaps 3;

QY 1 MLLFSVLISWVSTVGEGTLCDFPKIHGHGLYDEEDYNPFSQVPTGEVFFYSCYENFV 60
Db 1 MWLLSVLISRISSVGGEAMFCDFPKINHGLILYDEEKYKPPSQVPTGEVFFYSCYENFV 60

QY 61 SPKSFWRITCTEGSWPTPKCLRMCSFPFVKNGHSESSGLIHLEGGDTVQIICNTGYSL 120
Db 61 SPKSFWRITCTEGSWPTPKCLRMCSFPFVKNGHSESSGLIHLEGGDTVQIICNTGYSL 120

QY 121 QNNEKNISCVRGWSTPTPCSTKGECHVPILAEANVDAQPKKESYKGVGLKFSCKNLI 180
Db 121 QNNEKNISCVRGWSTPTPCSTKGECHVPILAEANVDAQPKKESYKGVGLKFSCKNLI 180

QY 121 QNNEKNISCVRGWSTPTPCSTKGECHVPILAEANVDAQPKKESYKGVGLKFSCKNLI 180
Db 121 QNNEKNISCVRGWSTPTPCSTKGECHVPILAEANVDAQPKKESYKGVGLKFSCKNLI 180
```

```
QY 181 RVGSDSVQCYQGWSPNFPCTCKGQVRSCTGPPPPQLSNGEVKEIRKEEYGHNEVVEYDCNPN 240
Db 141 ----- 140
QY 241 FIINGPKKIQCVGDGWTTLPTCVDQVCTGKIPELEYGVQVSPVPPYQHGVSVEVNCRNE 300
Db 141 ----- 140
QY 301 YAMIGNNMTICINGIWTLPVCMVATHOLKCRKCIAGVNIKTLLKLSGKBEHNHSRIYRCS 360
Db 141 ----- 140
QY 361 DIFRYRHSVCINGKWNPEVDCTEKREQFCPPPPQIPNAQNMNTTVNYQDGEKVAVLCKEN 420
Db 141 ----- 140
QY 421 YLLPEAKEIVCKDGRWQSLPRCVSTAYCGPPPSINNGDTSFPLSVYPPGSTVTVYRQCS 480
Db 141 -----RSTISAECGPPPIDNGDITSFLLSVYAPGSSVEYQCCN 180
QY 481 FYKLOGSVTVTCRNKQWSEPPRCPLDPCVVSENNMKNKNIQLKWRNDGKLYAKTGDAVEFO 540
Db 181 LYQLEGNRIITCRNQWSEPPKCLDPCVISQIMEKYNIKLNWTKNQKLYSRGTGDIIVEFV 240
QY 541 CK---FPHKAMISSPPFRAICQEGKFEYPIC 569
Db 241 CKSGYHPTK---SHSFRAMCQNGKLVYPSCE 268

RESULT 6
G35070
apolipoprotein H-related protein 13G1 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 27-Jul-1990 #sequence_revision 31-Dec-1993 #text_change 31-Mar-2000
C:Accession: G35070; A35071; B35071; H35070; I35070
R:Vik, D.P.; Munoz-Canoves, P.; Kozono, H.; Martin, L.G.; Tack, B.F.; Chaplin, D.D.
J. Biol. Chem. 265, 3193-3201, 1990
A:Title: Identification and sequence analysis of four complement factor H-related tra
A:Reference number: A35070; MUID:90153969; PMID:1689298
A:Accession: G35070
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-343 <VIR>
A:Cross-references: GB:M29008; NID:g192559; PIDN:AAA37414.1; PID:g309165; GB:J05259
A:Note: translation of the nucleotide sequence is not complete
C:Superfamily: apolipoprotein H; complement factor H repeat homology
C:Keywords: duplication
F:28-88/Domain: complement factor H repeat homology <FHL>
F:92-145/Domain: complement factor H repeat homology <FHR2>
F:154-208/Domain: complement factor H repeat homology <FHR3>
F:215-269/Domain: complement factor H repeat homology <FHR4>
F:273-334/Domain: complement factor H repeat homology <FHR5>

Query Match 28.5%; Score 906.5; DB 2; Length 343;
Best Local Similarity 32.9%; Pred. No. 9e-55;
Matches 187; Conservative 43; Mismatches 99; Indels 239; Gaps 5;

QY 1 MLLFSVLISWVSTVGEGTLCDFPKIHGHGLYDEEDYNPFSQVPTGEVFFYSCYENFV 60
Db 6 LLLLAIVLITSMFSAKGEVSLCDFPKIRHGLYDEKKNPEFSSVLSGLIYSCYENFA 65

QY 61 SPKSFWRITCTEGSWPTPKCLRMCSFPFVKNGHSESSGLIHLEGGDTVQIICNTGYSL 120
Db 66 SPNSFWRIITCTESGWSPTPKCLRLCFPPFVNGNSISGQTHVEGDIVQVVCNQYSL 125

QY 121 QNNEKNISCVRGWSTPTPCSTKGECHVPILAEANVDAQPKKESYKGVGLKFSCKNLI 180
Db 126 QNNQSTITCAEBSWITPKCISTN----- 149
QY 181 RVGSDSVQCYQGWSPNFPCTCKGQVRSCTGPPPPQLSNGEVKEIRKEEYGHNEVVEYDCNPN 240
Db 150 -----PTGK-----CGPPPIDNGDITSL----- 168
```

QY 241 FIIINGPKKIQCVGENTLPTCCEQVKTGCIPELEYGVQSPVPPYOHGVSVEVNCNE 300
 Db 169 -----SLPVASLSVEYQCQY 186
 QY 301 YAMIGNNMITCINGIWTLPVCMVATHQKCRCKIAGVNIKTLLKLSGKREFNHSRIRYRCS 360
 Db 187 YLLKGNKTIICRNGKWSPPFCI----- 209
 QY 361 DIFRYRHVSICNGKWNPEVDEKREOFCPPPPQIPNNAQNTMTTVNYODGEKVAVLCKEN 420
 Db 210 -----YPTGK----- 214
 QY 421 YLLPEAKEIVCKDGRWQSLPRCVSTAYCGPPPSINNGDTSFPLSVYPPGSTVYRQCS 480
 Db 215 -----CGPPPIDNGDITSLSLLEYPLSSVEYQOCN 246
 QY 481 FYKLGQSVTVTCRNKQWSEPPRCILDPCVSVSENNKNNIOLKWRNDGKLYAKTKDAVEFQ 540
 Db 247 YVVLGKGTITCRNGKWSPPFCI----- 306
 QY 541 CKFPHKAMISSPPFRAICQEGKFFYPIC 568
 Db 307 CKPRYKRAKGLSPRTQINGHINPTC 334

RESULT 7

S65551
 factor H - bovine (fragment)
 C:Species: Bos primigenius taurus (cattle)
 C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
 C:Accession: S65551
 R:Soames, C.J.; Day, A.J.; Sim, R.B.
 Biochem. J. 315, 523-531, 1996
 A:Title: Prediction from sequence comparisons of residues of factor H involved in the in
 A:Reference number: S65551; MUID:96202005; PMID:8615824
 A:Accession: S65551
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-669 (SOA)
 A:Cross-references: GB:X98697; NID:g1419423; PIDN:CAA67257.1; PID:g1419424
 C:Superfamily: complement factor H; complement factor H repeat homology
 F:55-114/Domain: complement factor H repeat homology <FHR1>
 F:296-349/Domain: complement factor H repeat homology <FHR2>
 F:355-412/Domain: complement factor H repeat homology <FHR3>
 F:416-471/Domain: complement factor H repeat homology <FHR4>
 F:476-530/Domain: complement factor H repeat homology <FHR5>
 F:538-592/Domain: complement factor H repeat homology <FHR6>
 F:599-651/Domain: complement factor H repeat homology <FHR7>

Query Match 27.0%; Score 861; DB 2; Length 669;
 Best Local Similarity 38.1%; Pred. No. 2.6e-51;
 Matches 164; Conservative 54; Mismatches 89; Indels 124; Gaps 5;

QY 23 CDRPKIHGFLYDEEDYNPFSQVPTGEVYVYSCYENVSFVSKSFWTRITCTEAGSWTPK 82
 Db 234 CSYPIVHKGRDLY--YSTRGFPPARVNOQFVYSDHHPVPPSRSDHDLACTAGSWSPPEP 291
 QY 83 CLRMCSPPFVKNGHSESSGLIHLEGDVQIICNTGYSLQNNKNSICVERGWSTPPIC-- 140
 Db 292 CLROCIFENYLENGHNRKREKYLOGETVRVHCYEGYSLQNDQNTMTCTESGWSPPPCIR 351
 QY 141 -----SFT----- 143
 Db 352 VKTCKSNIRIENGFLSESTFTYPLANKQTEYKCKPGVVTADGKTSGLITCLKNWSAQPV 411
 QY 144 ----- 143
 Db 412 CIKSCDRPVTEKARVKSDDGTWFRNLDRLDYECVDGVENRDRGRTGTSIVCGQDGSWKAAC 471
 QY 144 -KGECHVPPILEANVDAQPKKESYKVGDLVLFKSRKNLIRVGSQVQYQFGWSPNFPPTCK 202
 Db 472 YERECSIPMDPVLNAYPRKETVKVGDVLFKSSQGRIMVGDVQCYHFGWSPKLPCTCK 531

QY 203 -GQVRSRCPGPPQLSNGEVKEIRKEEYGHNEVVEYDCNPNFTINGPKKIQCVGDEWTLPT 261
 Db 532 VKVKSCALPPELPNGKRKEIKHEYAHNEVVEYACNPKFLMKGSHKIQCVDGESTALPV 591
 QY 262 CVEQVKTCGYIPELEYGVQSPVPPYOHGVSVEVNCNREYAMIGNNMITCINGIWTLP 321
 Db 592 CIEBERTIC-ELSDLDHGDVPSVPIIHGDSVSEFSCREAFMTIGRPFITCISGEMTQPPQ 650
 QY 322 CVATHQLKRCK 332
 Db 651 CIATDELKRCK 661

RESULT 8

A46013
 coagulation factor XIII subunit b - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 20-Jun-2000
 C:Accession: A46013
 R:Nonaka, M.; Matsuda, Y.; Shirolahi, T.; Moriwaki, K.; Nonaka, M.; Natsume-Sakai, S.
 Genomics 15, 535-542, 1993
 A:Title: Molecular cloning of the b subunit of mouse coagulation factor XIII and assi
 A:Reference number: A46013; MUID:93224141; PMID:8468048
 A:Accession: A46013
 A:Status: preliminary
 A:Molecule type: nucleic acid
 A:Residues: 1-668 (NON)
 A:Cross-references: GB:dl0071; NID:g303651; PIDN:BAA00963.1; PID:g303652
 A:Experimental source: liver
 A:Note: sequence extracted from NCBI backbone (NCBIN:129405, NCBIPI:129407)
 C:Superfamily: coagulation factor XIII chain b; complement factor H repeat homology
 F:25-87/Domain: complement factor H repeat homology <FH01>
 F:91-146/Domain: complement factor H repeat homology <FH02>
 F:153-208/Domain: complement factor H repeat homology <FH03>
 F:213-267/Domain: complement factor H repeat homology <FH04>
 F:274-327/Domain: complement factor H repeat homology <FH05>
 F:336-389/Domain: complement factor H repeat homology <FH06>
 F:396-450/Domain: complement factor H repeat homology <FH07>
 F:454-515/Domain: complement factor H repeat homology <FH08>
 F:524-578/Domain: complement factor H repeat homology <FH09>
 F:582-646/Domain: complement factor H repeat homology <FH10>

Query Match 24.6%; Score 783; DB 2; Length 668;
 Best Local Similarity 28.7%; Pred. No. 5.6e-46;
 Matches 187; Conservative 95; Mismatches 266; Indels 104; Gaps 18;

QY 7 VILISWVSTVGGEGTLCDFPKIHG---FLYD-EEDYNPFSQVPTGEVYVYSCYENFVS 61
 Db 9 ILLILSGELYAEKQCDFFTVENGRIAQYITFKSFYFPMS---VDKLSLFFCLAGYAT 65
 QY 62 PSKSFTRITCTEFGSPPTPKCLRMCSFPFVKNGHSESSGLIHLEGDVQIICNTGYSLQ 121
 Db 66 ESGKQEEQIRCTAEGWSPNRCYKCLKLPDLRNGYVSNQKLYKLOERNYSYCGSSGYKTT 125
 QY 122 --NNEKNNSICVERGWSTPPICSTFKGECVHPVILE----- 153
 Db 126 GKDDEEVHCLSGAGWSQSPSCRKEQETCLAPELEHGNYSTQRTFKVKDIAVATCTAGY 185
 QY 154 -----ANVDAQPKESYKVGDLVLFKSRKNL 179
 Db 186 TTTGKQTGEAEOAGNSWLTPOCNKLMCSRLRIENGFYFHPVKQTYEEDGLVQFFCHENY 245
 QY 180 IRVGSQVQYQFGWSPNFPPTCKGQVRSCOPPPQLSNGEVKEIRKEEYGHNEVVEYDCNP 239
 Db 246 YLSSDILQCYNGFWYSPESPICBGRNRNRCPPPPVPLNSKIQP-HSTYRHRGERVHIECEL 304
 QY 240 NFINGPKKIQCVGDEWTLPTCCEQVKTGCIPELEYGVQSPVPPYOHGVSVEVNC 297
 Db 305 NFVIQSEILLCEGKWKTEPPKCIEEKVACQSPQSPVNGVAHPHSEIYVYSGDKVTYRC 364
 QY 298 RNEYAMIGNNMITCINGIWTLPVCMVATHQKCRCK---IA-CVNITKTLKLSGKREFNH 352
 Db 365 GGSYLRSGSSTITCNRGRWTLPPCEV--ENIENCKRPPPPDIANGVVDGLL----ASYTTG 418

A:Cross-references: GDB:137201
C:Superfamily: complement C3d/Epstein-Barr virus receptor; complement factor H repeat homology
C:Keywords: glycoprotein
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-331/Product: complement factor H-related protein DOWN16 #status predicted <MAT>
F:23-83/Domain: complement factor H repeat homology <FH01>
F:87-141/Domain: complement factor H repeat homology <FH02>
F:147-204/Domain: complement factor H repeat homology <FH03>
F:211-265/Domain: complement factor H repeat homology <FH04>
F:269-330/Domain: complement factor H repeat homology <FH05>
F:108,186,206,310/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 22.38; Score 710; DB 2; Length 331;
Best Local Similarity 27.48; Pred. No. 2.5e-41;
Matches 156; Conservative 59; Mismatches 115; Indels 240; Gaps 8;

QY 1 MLLFSVLLSWSTVGGEGTCLDFPKIHGGFLYDEEDYNPFQVPTGEVFFYSCEYFV 60
DB 1 MLLINVLTLWVSCANGQVQKCDPDKHGGLFHEHMRPPVAVGKYISYCDHEFE 60

QY 61 SPKSFWRITCTEKGWSTPPKCLRMCSFPFVKNGHSESSGLIHLEGGDTVOIICNTGYSL 120
DB 61 TPGSYWDYIHTCTONGWSPVAVCLKCYFPYLENGYNGYRKFVQGNSTEVACHPGYGL 120

QY 121 QN-NEKNISCVRGWSTPPICSTFKGCHVPILAEANVDAQPKKESYKVGDLKFSCKNL 179
DB 121 PKVRQTITCTENGWSTPPIC-----IRDRTCSKSD 151

QY 180 IRVGSVSVQCYQFGWSPNFPTCKGQVRSCGPPQLSNGEVKEIRKEEYGHNEVVEYDCNP 239
DB 152 I-----ATADGN-----EIEGFISE-SSSIYILNKELQYKCRP 178

QY 240 NFIINGPKTKQCVGDEWTLPTCQEVQKTCGYIPELEYGVQVSPVPPYQHGVSVEVNCRN 299
DB 179 GY-----ATADGN-----EIEGFISE-SSSIYILNKELQYKCRP 178

QY 300 EYAMIGNNMTCTINGITWELPMCVATHQLKCKIAGVNIKTLKLSGKEFNHNSRIYRC 359
DB 187 -----SSGSITCLUR----- 195

QY 360 SDIFRYRHSVCINGKWNPEVDCTEKREQFCPPPPQIPNAQNTTTVNYQDGEKVAVLCKE 419
DB 196 -----NG----- 197

QY 420 NYLLPEAKEIVCKDGRWQSLPRCVSTAYCGPPPSINNGDTTSFPLSVYPPGSTVYRCQ 479
DB 198 -----WSAQPICINSSEKCGPPPIPSNGDTTSFLLKVVYVQSRVEYQCO 241

QY 480 SFYKLOGSVTVTCRNKQWSEPPRCLDPCVVSEENMKNKIQLKWRNDGKLYAKTGDVAF 539
DB 242 SYTELGSNVTCSNGWSEAPPRIHPCIIITENMKNKIQLKGRSDRKYAKTGDITIEF 301

QY 540 QCKFPFKAMISSPPFRAICQEGKFEYPICE 569
DB 302 MCKLGYNANTSILSFAVCREGIVEYPRCE 331

RESULT 11
A35068
complement factor H-related protein 3A4/5G4 - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 27-Jul-1990 #sequence_revision 31-Dec-1993 #text_change 20-Aug-1999
C:Accession: A35068; B35068; D35068; E35068; F35068; G35068
R:Wik, D.P.; Munoz-Canoves, P.; Kozono, H.; Martin, L.G.; Tack, B.F.; Chaplin, D.D.
J. Biol. Chem. 265, 3193-3201, 1990
A:Title: Identification and sequence analysis of four complement factor H-related transmembrane proteins from mouse
A:Reference number: A35070; MUID:90153969; PMID:1689298
A:Accession: A35068
A:Molecule type: mRNA
A:Residues: 1-452 <VIR>
A:Cross-references: GB:M29010; NID:g192561; PIDN:AAA37415.1; PID:g387128
A>Note: translation of the nucleotide sequence is not complete
C:Superfamily: complement factor H; complement factor H repeat homology

F:28-81/Domain: complement factor H repeat homology <FH1>
F:86-146/Domain: complement factor H repeat homology <FH2>
F:150-203/Domain: complement factor H repeat homology <FH3>
F:209-266/Domain: complement factor H repeat homology <FH4>
F:270-325/Domain: complement factor H repeat homology <FH5>
F:332-386/Domain: complement factor H repeat homology <FH6>
F:390-451/Domain: complement factor H repeat homology <FH7>

Query Match 20.3%; Score 646; DB 2; Length 452;
Best Local Similarity 24.9%; Pred. No. 8.5e-37;
Matches 157; Conservative 66; Mismatches 160; Indels 248; Gaps 13;

QY 1 MLLFSVLLSWSTVGGEG-----TL--CDFPKIHGGFLYDEEDYNPFQVPTGEVFFYSCEYFVSP 62
DB 6 MLLSNILLTAWLSTARGKEKTCSPPIYLGIVTPHRIHRSDEIRYECNMGYFYPVTGS 65

QY 21 -----TVSKCTPTGWIPIVPRCTLKPCEFPQYKGLRYEESLRPNFPYSIGNKYSYRCDNGFSPP 125
DB 66 TVSKCTPTGWIPIVPRCTLKPCEFPQYKGLRYEESLRPNFPYSIGNKYSYRCDNGFSPP 125

QY 63 SKSFWRITCTEKGWSTPPKCLRMCSFPFVKNGHSESSGLIHLEGGDTVOIICNTGYSLON 122
DB 126 SGYSWDYLRCTAGWEPVPCVRKCVFHYVENGDSAYWEKIYVQGSGLKQVCYNGSLON 185

QY 123 NEKNISCVRGWSTPPICSTFKGCHVPILAEANVDAQPKKESYKVGDLKFSCKNLIRV 182
DB 186 GDIMTCTENGWSPPKK----- 206

QY 183 GSDSVQCYQFGWSPNFPTCKGQVRSCGPPQLSNGEVKEIRKEEYGHNEVVEYDCNPFI 242
DB 207 -----KTCAS-----DIHIDNGFLSE--SSYIALNRETSYRCKQGYV 243

QY 243 INGPK---KIQCVDGDEWTLPTCQEVQKTCGYIPELEYGVQVSPVPPYQHGVSVEVNCRN 299
DB 244 TNGELSGSITCLONGWSPQSC---IKSC-----DMVPFENSIT--- 280

QY 300 EYAMIGNNMTCTINGITWELPMCVATHQLKCKIAGVNIKTLKLSGKEFNHNSRIYRC 359
DB 281 -----KNRTWFKL-----NDKLDYEC 297

QY 360 SDIF--RYRHSVCINGKWNPEVDCTEKREQFCPPPPQIPNAQNTTTVNYQDGEKVAVLC 417
DB 298 LVGFENEYKHT----- 308

QY 418 KENYLLPEAKEIVCKDGRWQSLPRCVSTAYCGPPPSINNGDTTSFPLSVYPPGSTVYRC 477
DB 309 -----KGSITCTYIGWSDTPSCYDSTRTCGPPPIDNGDITSLSLPEYPLSSVDYQ 360

QY 478 COSFYKLOGSVTVTCRNKQWSEPPRCLDPCVVSEENMKNKIQLKWRNDGKLYAKTGDV 537
DB 361 CQRYVLLKGNKTICTCRNGKWEPPPTCLHACVIPENIMEAHNIIKWRHTEKIYAHSGEDI 420

QY 538 EFQCKFPFKAMISSPPFRAICQEGKFEYPICE 568
DB 421 EFQCKRGYQKARGSUPPFRKTCINGTINYPTC 451

RESULT 12

S46199

Probable complement regulatory plasma protein SB1 - barred sand bass

C:Species: Paralabrax nebulifer

C:Date: 19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change 22-Mar-2002

C:Accession: S46199; S77894

R:Dahmen, A.; Kaidoh, T.; Zipfel, P.F.; Gligli, I.

Biochem. J. 301, 391-397, 1994

A:Title: Cloning and characterization of a cDNA representing a putative complement-re

A:Reference number: S46199; MUID:94318039; PMID:8042982

A:Accession: S46199

A:Molecule type: mRNA

A:Residues: 1-1053 <DAH1>

A:Cross-references: EMBL:L21703; NID:g639894; PIDN:AAA92556.1; PID:g639895
A:Experimental source: liver

A:Accession: S77894

THIS PAGE BLANK (USPTO)